



US009085767B2

(12) **United States Patent**  
Von Melchner et al.

(10) **Patent No.:** US 9,085,767 B2  
(45) **Date of Patent:** Jul. 21, 2015

(54) **ENHANCER-CONTAINING GENE TRAP VECTORS FOR RANDOM AND TARGETED GENE TRAPPING**

(75) Inventors: **Harald Von Melchner**, Kronberg/Taunus (DE); **Frank Schnütgen**, Alzenau (DE); **Wolfgang Wurst**, München (DE); **Patricia Ruiz**, Berlin (DE)

(73) Assignees: **FRANKGEN BIOTECHNOLOGIE AG**, Kronberg (DE); **GSF FORSCHUNGSZENTRUM FÜR UMWELT UND GESUNDHEIT GMBH**, Neuherberg (DE); **MPG MAX-PLANCK-GESELLSCHAFT ZUR FORDERUNG DER WISSENSCHAFTEN E.V.**, Munich (DE)

(\*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 1617 days.

(21) Appl. No.: **11/720,227**

(22) PCT Filed: **Nov. 28, 2005**

(86) PCT No.: **PCT/EP2005/056278**

§ 371 (c)(1),  
(2), (4) Date: **Aug. 3, 2007**

(87) PCT Pub. No.: **WO2006/056615**

PCT Pub. Date: **Jun. 1, 2006**

(65) **Prior Publication Data**

US 2010/0199360 A1 Aug. 5, 2010

(30) **Foreign Application Priority Data**

Nov. 26, 2004 (EP) ..... 04028194

(51) **Int. Cl.**

*C12N 15/86* (2006.01)  
*C12N 15/867* (2006.01)  
*C12N 15/65* (2006.01)  
*C12N 15/90* (2006.01)  
*C12N 15/10* (2006.01)

(52) **U.S. Cl.**

CPC .. *C12N 15/1051* (2013.01); *C12N 2740/13043* (2013.01); *C12N 2800/30* (2013.01); *C12N 2800/60* (2013.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2002/0194635 A1 \* 12/2002 Dunne et al. .... 800/15  
2004/0137490 A1 \* 7/2004 Finney et al. .... 435/6  
2004/0137572 A1 \* 7/2004 Finney et al. .... 435/69.1  
2004/0191873 A1 \* 9/2004 Young et al. .... 435/91.2  
2004/0221330 A1 \* 11/2004 Klimyuk et al. .... 800/278  
2005/0214793 A1 \* 9/2005 Lawn et al. .... 435/6  
2008/0104723 A1 \* 5/2008 Takeda et al. .... 800/25  
2013/0205421 A1 \* 8/2013 Petolino et al. .... 800/260

OTHER PUBLICATIONS

Wahlers et al, In Vivo Analysis of Retroviral Enhancer Mutations in Hematopoietic Cells: SPI/EGR1 and ETS/GATA Motifs Contribute to Long Terminal Repeat Specificity, Journal of Virology, Jan. 2002, pp. 303-312.\*

Baghdoyan et al, Capture of cytokine-responsive genes (NACA and RBM3) using a gene trap approach, Blood, Jun. 15, 2000, 95 (12), pp. 3750-3757.\*

Zambrowicz et al, Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells, Nature, vol. 392, Apr. 9, 1998, pp. 608-611.\*

De Palma et al, Blood, 2005, Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells, p. 2307-2315.\*

Groskreutz et al, Increased Expression and Convenience with the New pGL3 Luciferase Reporter Vectors, Promega Notes Magazine No. 50, 1995, p. 1-5.\*

\* cited by examiner

*Primary Examiner* — Maria Marvich

(74) *Attorney, Agent, or Firm* — Norris McLaughlin & Marcus, P.A.

(57) **ABSTRACT**

The present invention relates to a novel class of gene trap vector (enhanced gene trap vectors, eGTV) for efficiently identifying silent or weakly expressed target genes in mammalian genomes, methods of their production and methods for identifying and mutating target genes by using the enhanced gene trap vectors. The gene trap vectors of the present invention can also be used for inducing the expression of silent genes and enhancing the expression of weakly expressed genes. The use of the enhanced gene trap vectors for creating transgenic organisms to identify gene function and to validate pharmaceutical compounds prior to clinical applications is a further aspect of the present invention.

19 Claims, 5 Drawing Sheets

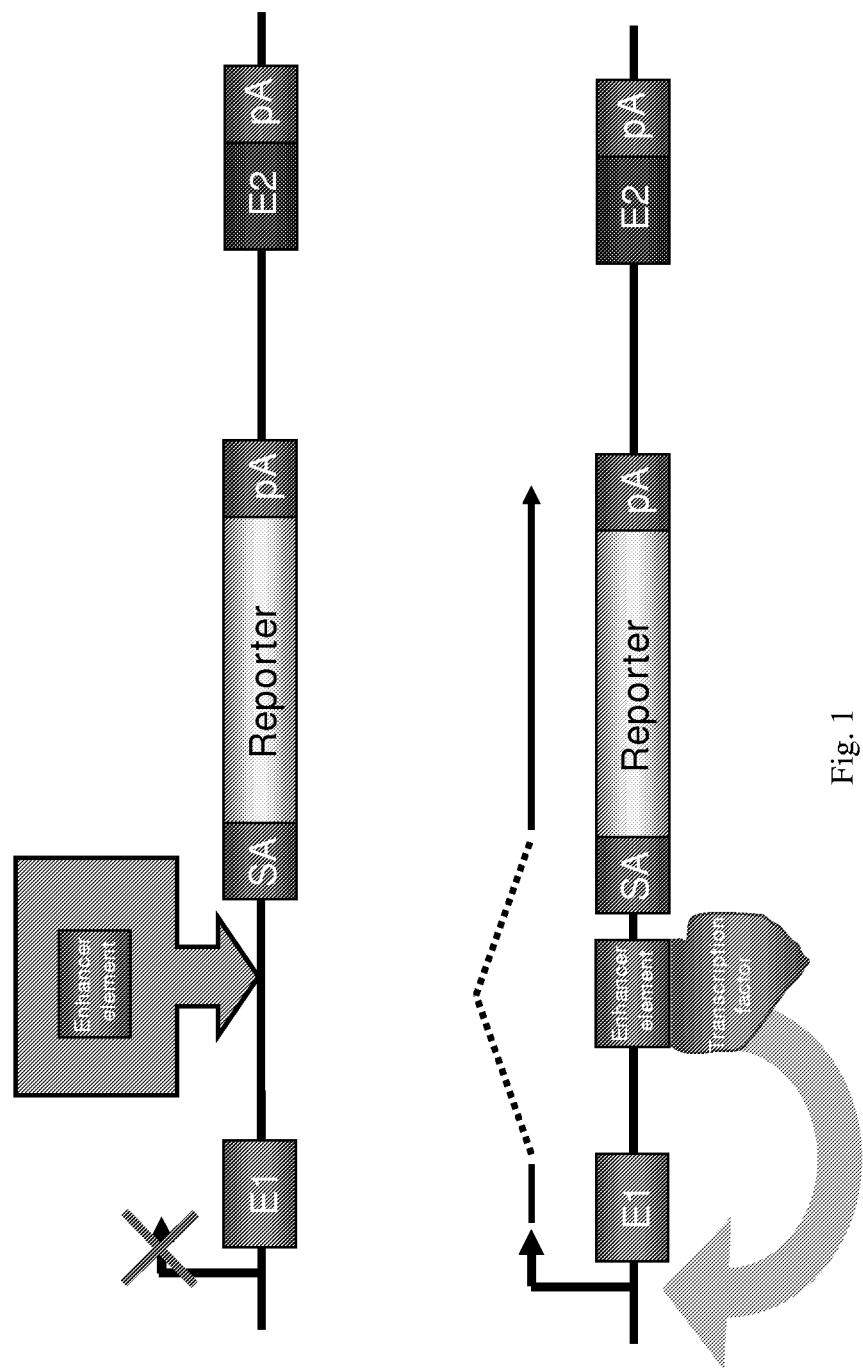
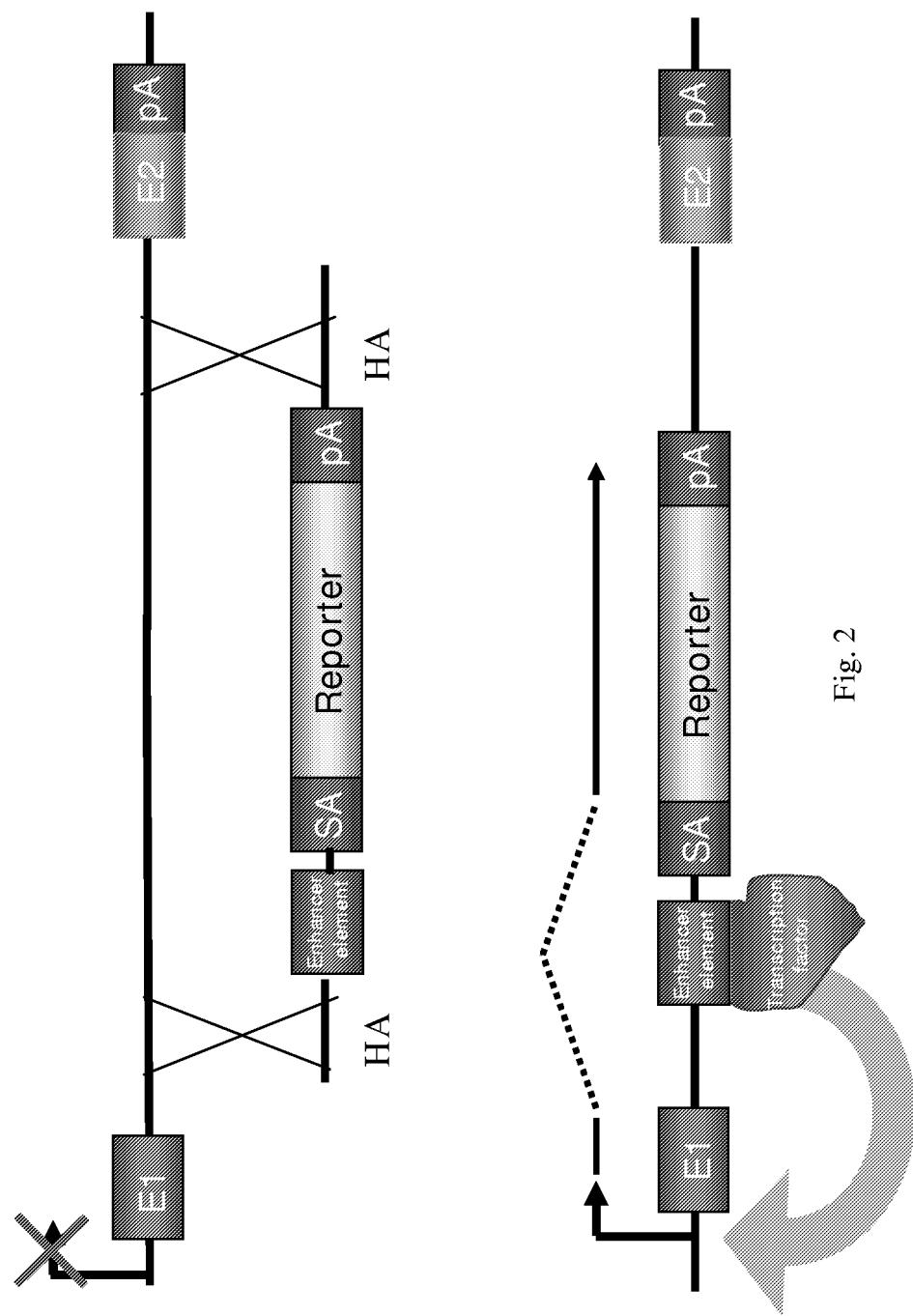


Fig. 1



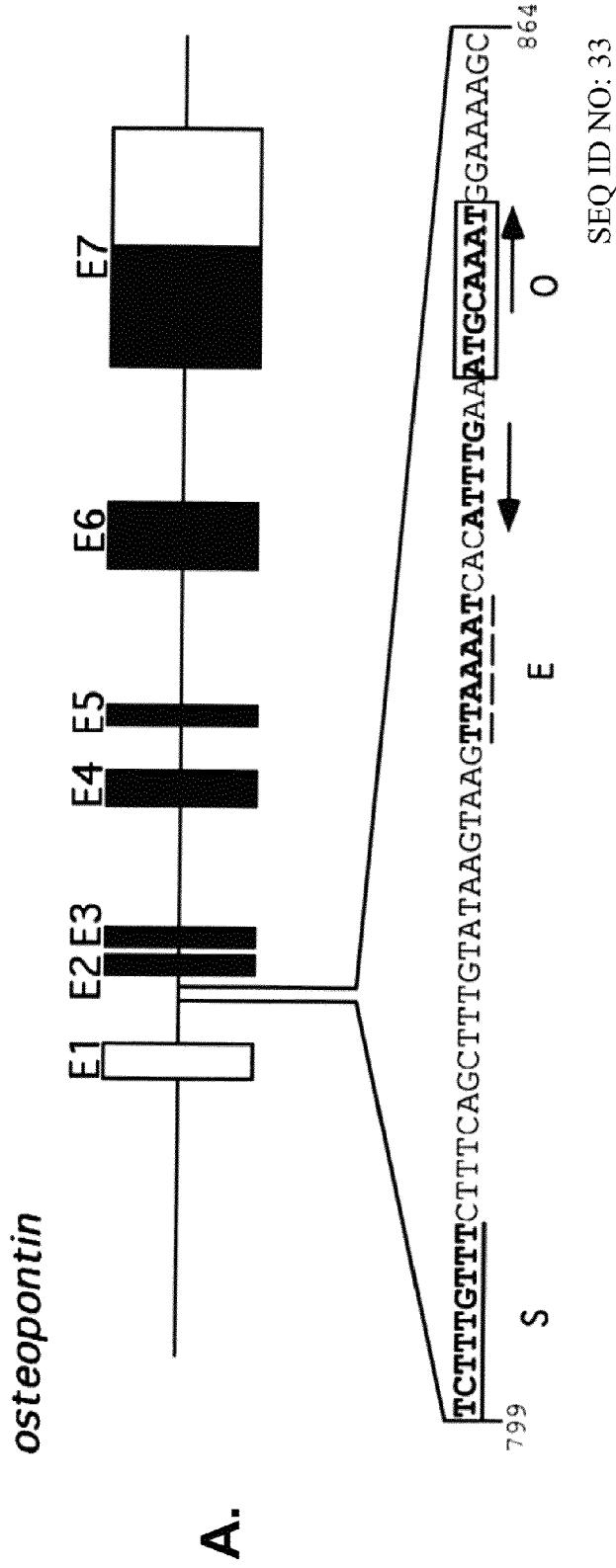


FIG. 3

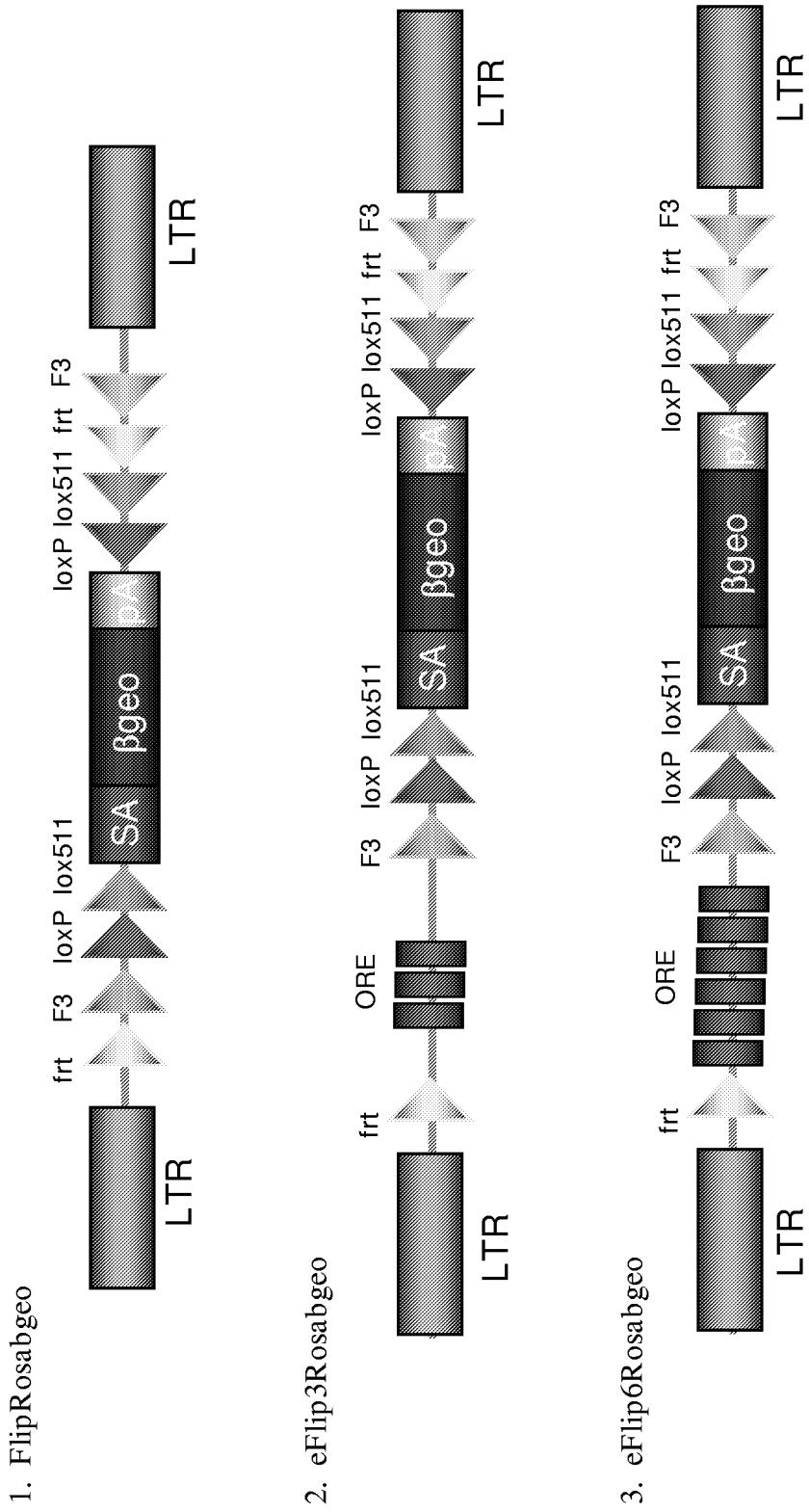


Fig. 4

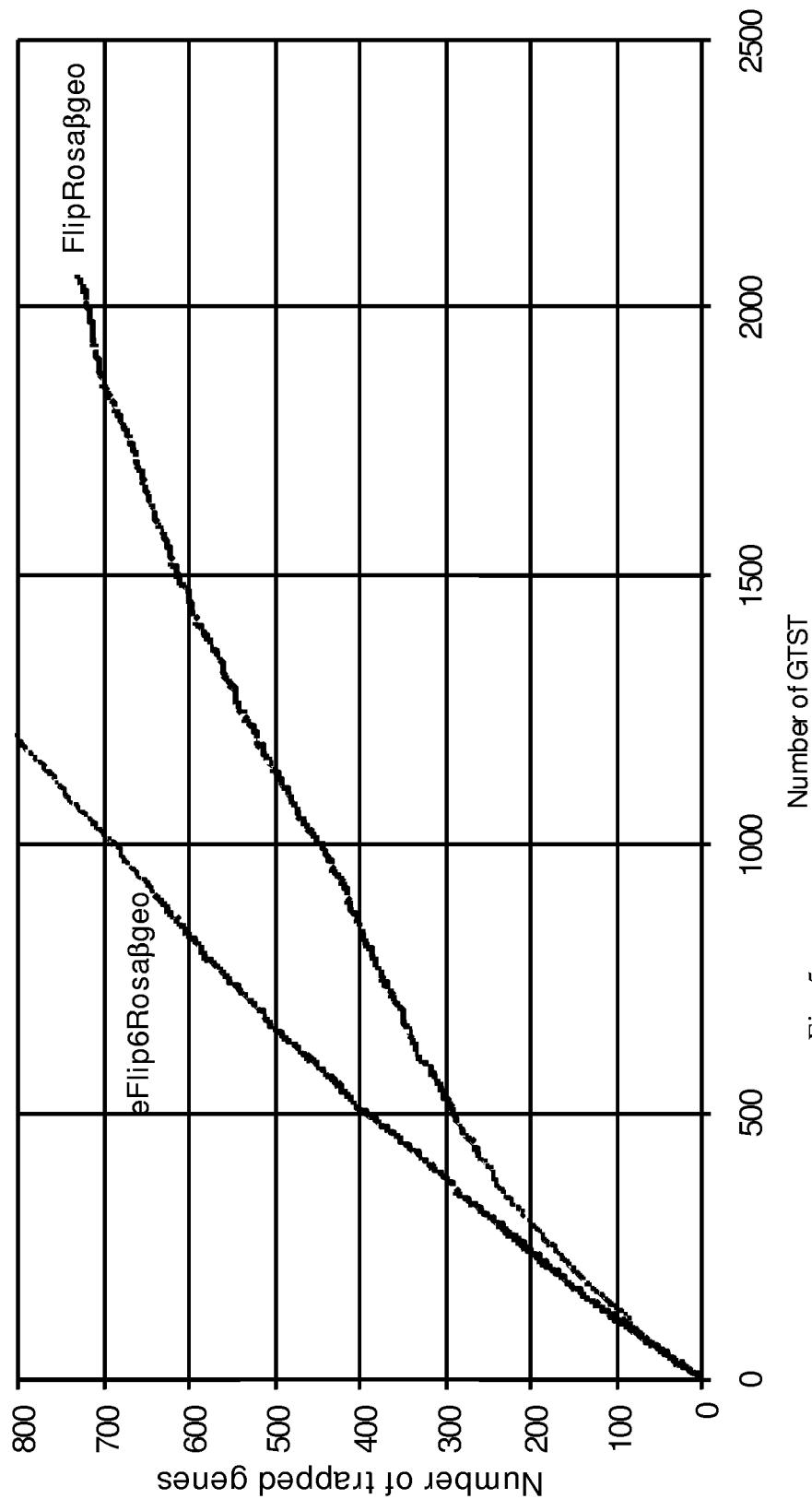


Fig. 5

1

**ENHANCER-CONTAINING GENE TRAP  
VECTORS FOR RANDOM AND TARGETED  
GENE TRAPPING**

The present invention relates to a novel class of gene trap vectors for efficiently identifying silent or weakly expressed target genes in mammalian genomes, methods of their production and methods for identifying and mutating target genes. The gene trap vectors of the present invention can also be used for inducing the expression of silent genes and enhancing the expression of weakly expressed genes. The use of the enhanced gene trap vectors for creating transgenic organisms to study gene function and to validate pharmaceutical compounds prior to clinical applications is a further aspect of the present invention. The novel gene trap vectors as referred to herein are termed as "enhanced gene trap vectors" (eGTV).

**BACKGROUND OF THE INVENTION**

In the past few years, a variety of gene trap vectors have been shown as being useful tools for the identification and analysis of permanently or transiently expressed genes. Standard gene trap vectors are DNA or retroviral vectors that insert a promoterless reporter gene into a large number of chromosomal sites. A classic gene trap vector integrates into introns, which are the non-expressed regions of a gene. Introns are flanked by exons, which are the expressed regions of a gene. Transcription of a trapped mammalian gene yields a primary messenger RNA consisting of exon, intron and vector sequences. Primary mRNA processing removes the intron sequences and splices the exons together at specific sites (splice sites) located at the 5' and 3' ends of each exon. As a result, the gene trap vector sequences encoding for the reporter gene become associated with the upstream exons in a processed fusion transcript from which a truncated cellular protein is translated together with the reporter protein.

With the completion of sequencing of the human and mouse genomes, the interest in tools suitable for performing genome-wide mutagenesis has significantly increased. Large scale insertional mutagenesis in mammalian cells has been most effectively induced with conventional gene trap vectors (Hansen, J. et al., Proc. Natl. Acad. Sci. USA 100:9918-22 (2003); Skarnes, W. C. et al., Nat. Genet. 36:543-4 (2004); Wiles, M. V. et al., Nat. Genet. 24:13-4 (2000); Zambrowicz, B. P. et al., Proc. Natl. Acad. Sci. USA 100:14109-14 (2003)). When selecting genes by means of their expression, recombinants will be obtained in which the reporter gene is fused to the regulatory elements of an endogenous gene. Transcripts generated by these gene fusion faithfully reflect the activity of individual cellular genes and serve as molecular tags to identify and/or clone any genes linked to specific functions. Thus, gene trap vectors simultaneously mutate and report on the expression of an endogenous gene at the site of insertion and provide a DNA tag for a rapid identification of the disrupted gene. The application of this technique in a genome-wide manner should allow for the identification of most, if not all, active transcripts in a genome and is thus an important tool for genome annotation. More importantly, gene trapping in mouse embryonic stem (ES) cells enables the establishment of ES cell libraries with mutations in a substantial fraction of genes in the mouse genome, which can be used to produce transgenic mice<sup>24</sup>. Thus, the gene trapping methodology enables the analysis of gene function in the context of an entire organism.

For some years targeted mutagenesis in pluripotent mouse embryonic stem (ES) cells has been used to inactivate genes

2

for which cloned sequences were available (Capecci, M. R., Trends Genet. 5:70-6 (1989)). Since ES cells can pass mutations induced *in vitro* to transgenic offspring *in vivo*, it is possible to analyze the consequences of gene disruptions in the context of entire organisms. As a result, numerous mouse strains with functionally inactivated genes ("knock out mice") have been created by this technology. However, targeted mutagenesis requires detailed knowledge of gene structure and organization as well as its physical isolation in a cloning vector. Overall, the generation of mutant mouse strains by this procedure is still time consuming, labor intensive, expensive and inefficient because it can handle only one gene at the time.

The principal element of a standard gene trap vector is a gene disruption and selection cassette (GDSC) consisting of a promoterless reporter gene and/or selectable marker gene flanked by an upstream 3' splice site (splice acceptor; SA) and a downstream transcriptional termination sequence (polyadenylation sequence; polyA; see FIG. 1). The GDSC is inserted into an intron of a target gene and transcription takes place from the upstream target gene promoter. Since the 3' end of the exon upstream of the vector insertion is flanked by a splice donor (SD) site, it is spliced to the GDSC resulting in a fusion transcript in which the upstream exons of the trapped gene are fused in frame to the reporter and/or selectable marker gene. Due to the presence of a polyA sequence in the GDSC, transcription is terminated prematurely, and, as a result, any exon(s) downstream of the GDSC are not transcribed anymore. Consequently, the processed fusion transcript encodes a truncated form of the target gene, consisting of the upstream exon(s), and the reporter/selectable marker gene.

From the above it becomes apparent that standard gene trap vectors can only disrupt genes that are actively transcribed in the target cell. Genes that are not expressed or expressed only too weakly for detection, i.e. at low expression levels, cannot be recovered by standard gene trapping. This poses a significant problem for genome-wide mutagenesis programs seeking a large scale and cost-effective functional analysis of the ~30,000 mammalian genes. In mouse embryonic stem (ES) cells, for example, only about one half of all genes are expressed, leaving ~15,000 genes inaccessible to standard gene trapping. The overall impact of a gene trap resource for elucidating gene function *in vivo* will thus rest on the fraction of the genome that is accessible with the standard gene trapping technology.

In order to trap genes that are not accessible to standard trapping, gene trap vectors that can be activated independently of gene expression have been developed previously. These vectors are based on a selectable marker gene flanked upstream by a constitutive promoter and downstream by a 5' splice site (splice donor, SD) (Zambrowicz, B. P. et al., Nature 392:608-11 (1998)). These elements are inserted downstream of a standard GDSC such as described above.

An insertion of these standard vectors into an intron of a gene induces splicing of the selectable marker gene, which, in turn, becomes associated with the downstream exon(s) of that gene. As a result, the cells express a fusion transcript initiating at the constitutive promoter and terminating at the polyA site of the trapped gene (=polyA trap). Since the selectable marker gene is expressed independently of the trapped gene's expression, poly-A traps should, at least in principle, enable the recovery of mutations in any gene.

However, there are some major drawbacks with these gene trap vectors and gene trapping methods. Several large scale screening efforts in ES cells with this technology have shown that polyA-containing gene trap vectors generate a high number of false positive recombinants and, more importantly, are

not considered to be highly mutagenic (Zambrowicz, B. P. et al., Proc. Natl. Acad. Sci. USA 100: 14109-14 (2003)). So far two main reasons have been cited for their poor performance: (i) the vectors frequently acquire cryptic polyA sites on the non-coding strands of genes, and (ii) selection is biased for gene trap insertions close to the 3' ends of genes, which are frequently non-mutagenic.

From the above it follows that there exists a need for gene trap vectors and gene trapping methods that overcome the above drawbacks, and which are efficient in the identification and mutation of cellular genes that are either not expressed or expressed too weakly to be detected by standard detection methodology. Thus, the provision of a gene trap strategy making most, if not all, genes of a genome accessible to effective trapping in a target cell would be highly desirable.

The problem underlying the present invention can thus be regarded as the provision of a gene trap vector and a gene targeting cassette that allows for the identification of gene products that are normally not expressed or expressed at non-detectable expression levels in a mammalian target cells. The solution provided by the present invention thus concerns a gene trap vector (eGTV) as defined in independent claim 1.

#### SUMMARY OF THE INVENTION

The present invention relates to a novel class of gene trap vectors that are capable of targeting genes independently of their expression.

In a first aspect of the invention, the gene trap vector of the invention comprises a gene disruption and selection cassette (GDSC) and at least one cell type-specific enhancer element that can be placed at any site within the vector, preferably upstream of the GDSC.

In a further aspect of the invention ubiquitous enhancer elements may be used. Such ubiquitous enhancer elements are e.g. the aldolase A enhancer (Moch C. et al., Gene Expr. 6:1-14 (1996)), the ployoma virus enhancer (Tanimoto K., et al., Nucleic Acids Res. 27:3130-3137 (1999)), the Oct-1 enhancer (Kemler I. et al., Nucleic Acids Res. 19:237-242 (1991)), and the murine adenosine desamidase enhancer (Winston J. H. et al., Somat. Cell Mol. Genet. 22:261-278 (1996)).

In a further aspect the gene trap vector of the invention comprises more than one enhancer element upstream and/or downstream of the GDSC. The GDSC of the invention comprises from 5' to 3': a splice acceptor sequence, a reporter gene and/or selectable marker gene and a transcription termination site. Preferably the transcription termination site comprises a polyA stretch consisting of adenylc acid (poly A) repeats.

In another aspect the enhancer elements of the present invention are located downstream of a target gene promoter and contain binding sites for transcription activating factors. Preferred enhancer elements of the invention are of the class of responsive elements containing repeat units of specific recognition sites for the corresponding enhancer element-binding partners.

In a further preferred embodiment, the enhancer elements are stretches of nucleic acid sequences of natural or synthetic, viral or non-viral origin that bind transcription activating factors in a sequence-specific manner. Examples are enhancer elements comprising hormone responsive elements, transcription factor binding elements, viral enhancer elements. Enhancer elements of the invention may comprise transcription factor binding sites for AP-1, AP-2, CRE, SRE, NF- $\kappa$ B, SRF, SP1, Oct1, Oct2, Oct3, Oct4 transcription factor binding sites. Preferably, the transcription factor binding sites are arranged as tandem repeats.

In a further aspect the enhanced gene trap vector of the invention comprises recombinase recognition elements for introducing GDSC inversions by site-specific recombinases. Examples of such recognition elements are FRT and IoxP recombination target sequences.

In a preferred embodiment, the gene trap vector of the invention contains one or more Oct-4 responsive enhancer elements that are inserted between two homotypic or heterotypic site-specific recombination targets (RTs).

10 In a preferred embodiment the reporter gene of the GDSC is  $\beta$ -galactosidase and the selectable marker gene is the neomycinphosphotransferase spliced together in a fusion gene.

In a preferred embodiment, the gene disruption and selection cassette (GDSC) and the enhancer element(s) are integrated in a retrovirus or a plasmid.

Preferred embodiments of retroviral gene trap vectors are FlipROSA $\beta$ geo, eFlip3ROSA $\beta$ geo and eFlip6ROSA $\beta$ geo comprising a puromycin resistance gene inserted downstream of the GDSC allowing for the quantification of gene trap insertions (WO 01/29208).

In another aspect, the present invention relates to a method for generating a transgenic non-human organism comprising

(i) incorporation of a gene trapping construct of the present invention into a cell of said non-human organism; and  
25 (ii) selection of cells in which said gene trapping construct is incorporated in a gene.

In a preferred embodiment of said method, it is suitable for identifying and/or isolating of a target gene in a non-human organism and comprises the steps:

(i) incorporation of a gene trapping construct according to the present invention in a vertebrate cell;  
30 (ii) selection of cells in which the gene trapping construct is incorporated in a gene; and optionally  
35 (iii) identification and/or isolation of the gene in which said gene trapping construct is incorporated.

In yet another aspect of the present invention, the gene trap vector can be used for mutating a target gene in a mammalian cell. The method for mutating a target gene in a mammalian cell, comprises

40 (i) transfection/infection of said cell with a gene trapping construct according to the present invention;  
(ii) incorporation of the gene trapping construct into the target gene, wherein the incorporation results in a truncated non-functional expression product.

45 It thus follows that the gene trap vectors of the present invention can be used for detecting, identifying or mutating a functional gene in a cell.

In another aspect, the gene trap vectors of the invention can be used for the generation of a gene trap library comprising gene trap insertions identified by the gene trapping methods of the present invention.

In a further aspect, the gene trap vectors of the invention can be used for targeted gene trapping in combination with homologous recombination.

55 In another aspect, the gene trap vectors of the invention can be used to create mouse mutant strains which are, among others, useful as models for genetic human disease and for validation of pharmaceutical compounds by monitoring in vivo effects of said compounds.

60

#### DESCRIPTION OF FIGURES

FIG. 1: Mechanism of eGTV activation. Insertion of enhancer elements binding a cell type specific transcription factor along with a gene disruption and selection cassette (GDSC) into an intron of a silent gene leads to its activation. SA=splice acceptor, pA=polyadenylation sequence.

FIG. 2: Enhanced target-directed gene trapping (targeted trapping). A GDSC is inserted together with enhancer elements binding a specific transcription factor into the intron of a silent target gene by homologous recombination. The eGTV activates gene expression and enables drug selection of homologous recombinants. SA=slice acceptor, pA=polyadenylation sequence, HA=homology arms.

FIG. 3: The Oct-4 intronic enhancer of the osteopontin gene. (A.) Position of the Oct-4 enhancer in the osteopontin gene. (B.) Sequence of the Oct-4 responsive element (ORE) used in the eGTV vectors. The Sox2 (S) binding sequence has been inactivated by mutagenesis. O=Oct-4 binding sequence, E=engrailed binding sequence.

FIG. 4: Conditional gene trap vectors. (Top) classic; (Bottom) enhanced. LTR=long terminal repeat, SA=slice acceptor,  $\beta$ geo= $\beta$ -galactosidase/neomycin-phosphotransferase fusion gene, pA=polyadenylation sequence, ORE=Oct-4 responsive elements. Triangles represent the homotypic and heterotypic site specific recombinase recognition targets for the FLPe (frt, F3) and Cre (loxP, lox511).

FIG. 5: Comparative rates of trapping exhibited by enhanced and non-enhanced gene trap vectors. Unique genes were identified by blasting (BLASTN) the GTSTs against the RefSeq database. The number of novel genes among accumulating insertions was determined chronologically.

#### DEFINITIONS

“Upstream” refers to nucleotide sequences that precede the codons specifying the mRNA or that precede (are on the 5' side of) the protein coding sequence.

“Downstream” refers to nucleotide sequences that follow the codons specifying the mRNA or that follow (are on the 3' side of) the protein coding sequence.

“Target Gene” defines a specific locus on a chromosome consisting of exons and introns to be trapped by a gene trap vector.

“Transcription activating factor” refers to a compound, e.g. a protein, polypeptide or peptide that has the ability of inducing gene expression by binding to a specific nucleic acid sequence (DNA, RNA).

“Transcription factor” refers to a compound, such as a protein, polypeptide or peptide that binds to one or more transcription factor recognition sites in the proximity of a gene promoter resulting in an activation of mRNA synthesis of said gene.

“Silent promoter” refers to a promoter that is not activated and consequently does not express the gene under control of this promoter.

“Silent gene” refers to a gene that is not expressed.

“Weakly expressed gene” refers to a gene, which is expressed at low expression level, wherein said expression levels are lower than that of a normally expressed gene. Also encompassed by the terms “weakly expressed” are a number of copies of the expression product that is too low to be detected by conventional standard detection methods.

“Gene disruption and selection cassette (GDSC)” refers to genetic elements comprising from 5' to 3' a splice acceptor sequence, a reporter and/or selection gene and a transcription termination site (e.g. poly A repeats).

“Gene trapping” refers to a random mutagenesis approach in functional genomics and is based on the random integration of a gene disruption and selection cassette into a genome.

“Targeted trapping” refers to a gene specific mutagenesis approach in functional genomics and is based on the insertion of a GDSC into the genome by homologous recombination.

“Gene trap vector” refers to a promoterless gene trapping construct consisting of nucleic acid, wherein the gene trapping construct is capable of generating a fusion transcript with the targeted endogenous gene. The presence of splice acceptor elements in the gene trap vector results in the generation of a fusion protein after its insertion into introns.

“Enhanced gene trap vector” (eGTV) refers to a gene trap vector, which contains one or more enhancer elements in addition to a gene disruption and selection cassette (GDSC).

“Reporter gene” refers to a gene encoding for a gene product (e.g. CAT,  $\beta$  galactosidase,  $\beta$ geo, GFP, EGFP, alkaline phosphatase) that can be readily detected by standard biochemical assays.

“Selectable marker gene” refers to a gene whose expression in a transfected cell allows for the isolation of gene trap vector-expressing cells in drug-containing media (e.g. neomycin, puromycin, diphtheria toxin).

“PolyA” (A=adenylic acid) refers to a nucleic acid sequence that comprises the AAUAAA consensus sequence, which enables polyadenylation of a processed transcript. In a gene disruption or selection cassette (GDSC), the polyA sequence is located downstream to the reporter and/or selectable marker gene and signals the end of the transcript to the RNA-polymerase.

“Splicing” refers to the process by which non-coding regions (introns) are removed from primary RNA transcripts to produce mature messenger RNA (mRNA) containing only exons.

“5' splice site” (splice donor SD) and “3' splice site” (splice acceptor SA) refer to intron flanking consensus sequences that mark the sites of splicing.

“Enhancer element” or “enhancer” refers to a nucleic acid sequence, which can increase the levels of transcription of a gene into messenger RNA. Typically, an enhancer element functions in either orientation and at various distances from a cellular promoter.

“Responsive enhancer element” refers to an enhancer element, which is specific for a particular transcription factor. When bound by the specific transcription factor, the levels of expression of the gene are enhanced. An example is the Oct-4 responsive element (ORE), which in combination with the transcription factor Oct-4 increases the gene's expression level.

“GDSC inversion” refers to an aberration in which a GDSC segment is deleted, reinserted and turned by 180 degrees from its original orientation, so that the gene sequence for the segment is reversed with respect to that of the rest of the chromosome. Said inversions can be accomplished by using recombinase enzymes (e.g. Cre, FLPe,  $\phi$ C31).

“Tandem repeats” refers to copies of genetic elements repeated one after another along a genomic or vector site.

“Homotypic” means being of the same type or form.

“Heterotypic” means being of different type or form.

“ROSA” (Reverse-Orientation-Splice-Acceptor) refers to a gene trap cassette inserted into a retroviral backbone in reverse transcriptional orientation relative to the retrovirus (Friedrich, G., Soriano, P., Genes Dev. 5:1513-1523 (1991)).

Sequence Listing	
SEQ ID NO:	free text
1	FlipROSA $\beta$ Geo(int)
2	pBABErf1
3	FlipROSA $\beta$ Geo
4	eFlip3ROSA $\beta$ Geo

Sequence Listing	
SEQ ID NO:	free text
5	eFlip6ROSAβGeo
6	FlipROSAβGeoPuro
7	eFlip3ROSAβGeoPuro
8	eFlip6ROSAβGeoPuro
9-16	primer II, I3, I2, I4, I6, I8, I5 and I7
17-18	oligonucleotides P5 and P6
19-20	primer P7and P8
21-24	primer
25-32	oligonucleotides I16, SR1, I15, ISR2, I14, iPCRu3, I13 and iPCRu4

## DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a novel class of gene trap vectors with the ability to trap silent or weakly expressed genes in a vertebrate genome. More specifically, the gene trap vectors and gene trapping constructs of the invention allow for trapping genes that are expressed at a low expression level, i.e. genes that usually escape detection by standard detection methods because the copy number of their expressed products is too low to be detectable by standard detection methodology (e.g. by utilization of antibodies in Western blotting or ELISA (enzyme-linked immunosorbent assay)). This novel class of gene trap vectors will be in the following referred to as "enhanced gene trap vectors" (eGTV).

The invention is based on the finding that the expression of a silent or weakly expressed gene can be induced or enhanced by using a gene trap vector construct that contains one or more enhancer elements capable of activating the trapped gene's promoter. This, in turn, activates gene trap expression, which enables further analysis of the gene. By using enhancer elements that are responsive to cell type specific transcription factors in gene trap vectors, it is possible to disrupt genes that are not accessible to standard methodology.

Gene trap vectors have been designed containing a standard GDSC in combination with at least one enhancer element that can insert either randomly (gene trapping) or specifically (targeted gene trapping) throughout the genome. It is preferred that the gene trap vectors of the invention integrate in non-expressed sites of the genome (introns), i.e. between the expressed regions (exons) of a gene. Following transcriptional activation of the trapped gene by means of the inserted enhancer elements, a fusion transcript is generated between the upstream exons and the GDSC resulting in a selectable mutation. To determine the identity of the trapped gene, the fusion transcript can be reverse transcribed, amplified by PCR and subsequently sequenced.

Conventional gene trap vectors and classic gene targeting methods require gene expression for the successful mutagenesis of target genes. However, genes that are not actively transcribed cannot be disrupted by these methods. Depending on the cell type, between 50-80% of all genes are not transcribed under natural conditions because the cell does either not require the gene product for survival and proliferation, or because the cell has ceased to differentiate any further.

We therefore sought to provide a gene trap vector with the ability to either induce gene expression or to enhance gene expression of a target gene. To achieve this, the gene trap vectors of the invention contain at least one enhancer element serving as a binding site for transcription activating molecules that, when bound, are able to turn on the transcription of the trapped gene.

The gene trap vectors of the invention are equipped with cell type specific enhancer elements that are placed into the target genes upon vector insertion. It is preferred that one, two or more enhancer elements are arranged within the gene trap vector depending on the degree of activation and/or class of genes to be analyzed. The enhancer elements can be placed at various distances from the promoter of the gene of interest, preferably they are arranged downstream of the promoter region. In preferred embodiments, the enhancer elements are of natural or synthetic, viral or non-viral origin and bind transcriptional activators. In yet another preferred embodiment of the invention, the enhancer elements comprise responsive elements such as hormone responsive elements, transcription factor binding elements, and viral enhancer elements. The specific selection of the enhancer elements will depend on the cell type and class of genes to be analyzed. In a further preferred embodiment, the enhancer elements comprise binding sites of the following transcription factors: AP-1, AP-2, CRE, SRE, NF-κB, SRF, SPI, Oct1, Oct2, Oct3, Oct4 (Nakabeppu, Y. et al., Cell 55:907-15 (1988); Bosher, J. M. et al., Oncogene 13:1707-7 (1996); Gotquin, V. et al., Genes Dev. 12 2073-90 (1998); Scholer, H. R. et al., Nature 344:435-9 (1990)). Preferably, the transcription factor binding sites are arranged as tandem repeats.

The binding of a transcription factor to its cognate enhancer element induces the expression of silent genes and enhances the expression of weakly expressed genes. By activating gene expression, the enhancer-bound transcription factors also activate the inserted enhanced gene trap vector (eGTV) of the invention in much the same way than standard gene trap cassettes are activated, however, with the advantages referred to below. Thus, by using the enhanced gene trap vectors of the invention, it is possible to identify and select genes that would normally escape such a selection (see FIG. 1).

The enhanced gene trap vectors of the present invention are not only suitable for random mutagenesis but can also be used for targeted gene trapping involving the introduction of a GDSC along with one or more specific enhancer elements into a silent gene by homologous recombination (see FIG. 2). The activation of a target gene by specifically binding a transcription factor simultaneously activates the GDSC, and thus enables the recovery of homologous recombinants by drug selection. The number of genes accessible to trapping in a particular target cell is thereby increased above the number of genes accessible to standard trapping (see examples).

In a preferred embodiment, these enhancer elements include intronic Oct-4 enhancer elements derived from the osteopontin gene (see FIG. 3). Oct-4 is a transcription factor, which is highly and specifically expressed in undifferentiated embryonic stem (ES) cells (Scholer, H. R. et al., Nature 344:435-9 (1990)). The Oct-4 transcription factor specifically recognizes and binds to the conserved octamer motif -ATG-CAAAT-, which is present in Oct-4 enhancers (see FIG. 3). In ES cells, insertions of a gene trap vector of the invention comprising Oct-4 responsive enhancer elements (OREs) into the introns of silent or weakly expressed genes will activate these genes by binding the Oct-4 transcription factor (see examples). This, in turn, induces GDSC expression, which enables selection for gene trap events.

It is preferred that the reporter gene of the invention is γ-galactosidase, βgeo or a gene encoding for a luminescent gene, e.g. a fluorescent gene such as green fluorescent protein (GFP) or firefly luciferase. It is further preferred that the selectable marker gene encodes an anti-antibiotic such as neomycin- and hygromycin-phosphotransferases or puromycin-acetyl-transferase.

In another aspect of the invention, the gene trap vectors comprise target sequences for site-specific recombinases to enable postinsertional GDSC inversions and/or excisions. Thus, in an embodiment the enhanced gene trap vectors of this aspect of the invention contains: (i) a GDSC consisting of a β-galactosidase-neomycinphosphotransferase fusion gene (βgeo) flanked by SA- and polyA-sequences for gene trapping, (ii) elements enabling GDSC inversions by site specific recombinases for conditional mutagenesis, and (iii) one or more Oct-4 responsive enhancer elements (ORE), preferably inserted between two heterotypic site-specific recombination targets (RTs) (see FIG. 3).

Any suitable recombinase can be used for inversion, such as Cre recombinase, Flp recombinase or φC31 recombinase.

It is preferred that the gene trap vectors of the invention are contained in a retrovirus or a plasmid. These gene trapping constructs contain in addition to a GDSC at least one enhancer element for transcriptional activation of an upstream or downstream gene promoter. Retroviruses insert a single copy of the reporter and/or selectable marker cassette per locus, with no rearrangement of flanking sequences. In particular, retroviruses have a preference for insertions at the 5' ends of genes.

The gene trap vectors of the invention are preferably retroviral gene trap vectors selected from the group consisting of FlipROSAβgeo, eFlip3ROSAβgeo and eFlip6ROSAβgeo (WO 01/29208). These vectors preferably comprise a puromycin resistance gene inserted downstream of the GDSC allowing for the quantification of most if not all gene trap insertions.

The above specified retroviral or plasmid based gene trap vectors are suitable for disrupting both silent and expressed genes across any mammalian genome (e.g. human, mouse). For example, by using the gene trap vectors of the invention in a genome wide manner, a large collection of embryonic stem (ES) cell lines harboring gene trap insertions in single genes can be assembled and used to make mutant mice. In particular, for pharmaceutical research seeking to validate the utility of specific genes and their products as targets for drug development, mutant mice are excellent genetic tools.

In another aspect, the invention relates to a method for the identification of a functional yet silent gene in mammalian cells. The method comprises the transduction of cells with an enhanced gene trap vector as described herein and the incorporation of the reporter gene and/or selectable marker cassette into genomic sites. If the vector inserts into the intron of silent genes, gene trap vector-induced promoter activation results in GDSC expression, which in turn enables cell selection. Preferably, the disrupted gene is identified by RT-PCR (RACE) or PCR (PCR=polymerase chain reaction; RT=reverse transcription) (Hansen, J. et al., Proc. Natl. Acad. Sci. USA 100:9918-22 (2003); von Melchner et al., Proc. Natl. Acad. Sci. USA 87:3733-7 (1990)).

In a preferred embodiment, the method of the present invention is adapted to reach a cost-effective saturation of the genome with insertional mutations in the fastest possible way. The steps involved in this method are: (i) transducing a large number of embryonic stem (ES) cells with enhanced gene trap vectors (eGTV) of the invention, (ii) selecting eGTV-expressing clones and establishing cell lines from them, (iii) creating an ES cell bank containing ES cell lines with mutations in single genes, (iv) amplifying genomic sequences adjacent to the eGTV insertion by PCR and/or sequences appended to eGTV transcripts by RT-PCR (RACE) from the ES cell lines, (v) sequencing the amplification products to obtain cell line specific gene trap sequence tags (GTSTs, "flank bank"), (vi) identifying and cataloguing the disrupted

genes by GTST homology searches in the public databases, (vii) making mutant strains of mice using ES cells from the ES cell bank.

The invention further encompasses a method for mutating a functional gene within the genome. The mutation is introduced by incorporating a gene trap vector of the invention into intronic sites of a gene. As a result of the splicing process (see supra), a fusion product will be obtained comprising one or more exons of the gene and the reporter/selectable marker gene cassette. Due to the fact that transcription terminates at the polyA site, the downstream exons of the gene will not be part of the fusion product. Therefore, the mRNA of the gene is not complete but truncated since every exon that follows the gene trap vector will not be expressed.

A further aspect of the invention relates to a method for producing both "null" and "conditional" mutations in genes of an organism regardless of whether the genes are expressed or not. In a preferred embodiment the method comprises the following steps:

- 15 20 (i) incorporation of a gene trapping vector construct in a suitable cell;
- (ii) selection of cells having the vector construct incorporated in a gene by expression analysis of the selectable marker;
- (iii) identification and/or isolation of the gene in which the construct is incorporated by PCR or Reverse Transcriptase (RT)-PCR.

The gene trap vectors of the invention and their uses allow both random or targeted mutagenesis in mammalian cells. For a targeted-insertional mutagenesis (=targeted gene trapping), 30 sequence information of the gene of interest is required for specific integration, whereas in a random insertional mutagenesis the integration occurs at non-specific sites. A method for targeted-insertional mutagenesis using enhanced targeted gene trapping constructs of the present invention 35 comprises the following steps:

- 35 40 (i) design of a gene trapping construct comprising the enhancer elements and the GDSC from an enhanced gene trap vector flanked by sequences homologous to the an intron preferably flanking a 5' exon of the specific target gene (=homology arms; see FIG. 2);
- (ii) transduction of the gene trapping construct into a suitable cell;
- (iii) isolation of homologous recombinants by selecting for GDSC expression;
- 45 (iv) verification of homologous recombination in the selected cells by a suitable detection method, preferably 5' RACE, genomic PCR or Southern blot analysis.

Taken together the enhanced gene trap vectors (eGTV) of the invention and the methods employing them are suitable to solve the drawbacks of conventional gene trap- and gene targeting vectors currently used in the field. In particular, it shows that the eGTV have the capability to

- 50 (i) induce mutations in genes regardless of their expression, including genes for which cloned sequences are not available;
- (ii) increase the number of genes accessible to trapping and targeted trapping in comparison to the number presently accessible by standard gene trap and targeted trapping vectors;
- 55 (iii) increase the gene trapping rate by reducing redundancy, and thereby saving costs.

In its application to pharmaceutical research, the eGTV technology greatly assists the creation of a library of ES cell clones, which includes all genes of the mouse genome (~30,000), modified by the insertion of an eGTV vector. Sequence analysis from the modified alleles allows, as described above, to identify the genes modified in individual clones, resulting

11

in a database. This collection is the resource required to quickly generate a large number of mutant mouse strains for pharmaceutical research. Such mutant strains are an optimal tool to study the function of genes in the mouse as a mammalian model organism for human disease. The evaluation of gene function, in particular in mouse strains prone to develop a disease, allows to validate the utility of an individual gene for pharmaceutical drug development. For example, a given number of genes could be thought to be involved into insulin receptor signaling as a therapeutic target for Diabetes treatment. The generation and physiological analysis of mouse mutants for these genes would identify which gene products play a role in insulin signaling. Subsequently, drug development could be focused only on such "validated targets". Thus, within pharmaceutical drug development, mutants serve as a validation instrument to identify useful target molecules.

In addition, a subset of mutants could develop diseases, which mimic known human disorders and serve as a model for their treatment.

Once validated targets are available, the mutants can be used in combination with gene expression profiling to determine "on" and "off" target effects of candidate drugs.

Since the use of the gene trap technology is not restricted to murine ES cells and mouse mutants its application can be extended to any other vertebrate or invertebrate model organism (e.g. rats, zebra fish, *Drosophila*) to characterize the biological function of selected genes. As described above for the mouse, such mutants could be used for the validation of target genes for pharmaceutical drug development.

In another application the gene trap technology can be used to validate the utility of selected genes of plants for agricultural purposes. Plant mutants can be used to identify valuable target genes for herbicide development as well as to identify genes involved into the fertility of economically useful species.

As further exemplified in the examples below, the eGTV vectors of the invention do not only allow the identification of expressed genes in a cell but also the identification of yet unidentified, silent or poorly expressed genes. In summary, eGTVs provide superior tools for the field of genomics and functional genetical analysis.

In the following examples, material and methods of the present invention are provided. It should be understood that these examples are for illustrative purpose only and are not to be construed as limiting this invention in any manner. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entity for all purposes.

## EXAMPLES

### Example 1

To test the trapping efficiency of enhanced gene trap vectors, the retroviral gene trap vectors FlipROSA $\beta$ geo, eFlip3ROSA $\beta$ geo and eFlip6ROSA $\beta$ geo (FIG. 4) were equipped with a puromycin resistance gene, which is independently transcribed from a pgk promoter. This element, inserted downstream of a  $\beta$ -galactosidase (LacZ)-neomycin-phosphotransferase fusion gene ( $\beta$ geo) flanked by SA- and polyA-sequences enables the quantification of all gene trap insertions across the genome of a target cell. Enhanced gene trap vectors contain in addition to a GDSC either 3 or 6 Oct-4 responsive elements (OREs) arranged in tandem repeats (see FIG. 4).

12

A. Construction of the Gene Trap Vectors FlipROSA $\beta$ geo, eFlip3ROSA $\beta$ geo and eFlip6ROSA $\beta$ geo.

Construction of the gene trap vector FlipROSA $\beta$ geo: For the construction of FlipROSA $\beta$ geo an overlap extension PCR strategy was chosen using the oligonucleotides I1 (5'-CGC CTC CTC TTC CTC CAT CC-3'; SEQ ID NO:9) and I3 (5'-ACT CTT CCG CTT CCT CGC TCC ACC GCG GCT TCG AGA CCG T-3'; SEQ ID NO:10) for amplification of the 5' f3-frt recombinase target (RT) sequences from FlipROS-  
10 A $\beta$ geo(int) (see SEQ ID NO:1) and the oligonucleotides I2 (5'-GGG CCT CTT CGC TAT TAC GC-3'; SEQ ID NO:11) and I4 (5'-ACG GTCTCG AAG CCG CGG TGG AGC GAG GAA GCG GAA GAG T-3'; SEQ ID NO:12) for amplification of the 5' Iox511-IoxP RTs from pFIEx; Schnutgen, F. et al., Nat. Biotechnol. (2003)). The two PCR products were purified, annealed and re-amplified using the oligonucleotides I1 and I2. The resulting PCR product was cloned into the BamHI site of the retroviral vector pBABESrfl (modified from pBABE puro; see SEQ ID NO:2) (Morgenstern, J. P.,  
20 Land, H., Nucl. Acids. Res. 18:3587-3596 (1990)) to obtain -pBLF5'. Similarly, the 3' RTs were amplified with the oligonucleotides I6 (5'-GCT CCT CGC CCT TGC TCA CC-3'; SEQ ID NO:13) and I8 (5'-TAG AAG CGG TTT TCG GGA GAA TAC GAC TCA CTA TAG GGC G-3'; SEQ ID NO:14)  
25 for IoxP-Iox511 and with the oligonucleotides I5 (5'-TGC TGG CCT TTT GCT CAC AT-3'; SEQ ID NO:15) and I7 (5'-CGC CCT ATA GTG AGT CGT AIT CTC CCG AAA ACC GCT TCT A-3'; SEQ ID NO:16) for the frt-f3. The two products were purified, annealed and re-amplified with the  
30 oligonucleotides I6 and I7. The resulting PCR product containing all four RTs was cloned into the EcoRI restriction site of pBLF5' to obtain pBLF. The GDSC (SA- $\beta$ geo-pA) was recovered as an XbaI restriction fragment from ROSA $\beta$ geo (Friedrich, G., Soriano, P., Genes. Dev. 5:1513-1523 (1991))  
35 and cloned into the SnaBI restriction site of pBLF to obtain the final gene trap vector -FlipROSA $\beta$ geo-. The final vector was verified by DNA sequencing (see SEQ ID NO:3).

Construction of the gene trap vector eFlip3ROSA $\beta$ geo: Oct 4 responsive elements (ORE) from the osteopontin gene (Botquin, V. et al., Genes Dev. 12:2073-90 (1998) were obtained by annealing the oligonucleotides P5 (5'-GAT CCT GCA CTG ACC TTT CAG CTT TGT ATA ATG TAA GTT AAA ATC ACA TTT GAA ATG CAA ATG GAA AAG CA-3'; SEQ ID NO:17) and P6 (5'-GAT CTG CTT TTC CAT TTG CAT TTC AAA TGT GAT TTT AAC TTA CAT TAT ACA AAG CTG AAA GGT CAG TGC AG-3'; SEQ ID NO:18) and cloning into the BamHI/BglII sites of pEGFP-N1 resulting in pEGFP-O1. pEGFP-O2 to pEGFP-O6 was obtained by subsequent ligation of additional to pEGFP-O1.  
40 3-6 OREs obtained from pEGFP-O3 and pEGFP-O6 were cloned as BamHI/BglII fragments into the BglII site located in the spacer between the frt and the f3 sites of FlipRosa $\beta$ geo to obtain eFlip3ROSA $\beta$ geo and eFlip6ROSA $\beta$ geo. The primary structure of the final vectors was confirmed by DNA  
45 sequencing (SEQ ID NOs:4 and 5).

B. Construction of the Gene Trap Vectors FlipROSA $\beta$ geoPuro, eFIip3ROSA $\beta$ geoPuro and eFIip6ROSA $\beta$ geoPuro

Construction of the gene trap vector FlipROSA $\beta$ geoPuro: The puromycin cDNA was obtained from pBabePuro (Morgenstern, J. P., Land, H., Nucl. Acids. Res. 18:3587-3596 (1990)) by PCR using the primers P7 (5'-GGG GGC TGC AGA CTT ACA GCG GAT CCC CTC AGG CAC CGG GCT TGC-3'; SEQ ID NO:19) and P8 (5'-GGG GGC TGC AGC CAA TAT GAC CGA GTA CAA GCC CAC-3'; SEQ ID NO:20). The puromycin cDNA was then used to replace the neomycin resistance gene of expression plasmid pd383 (Zazopoulos, E. et al., Nature 390:311-5 (1997)), obtain  
50

pD383/puro. The pgk promoter-puro-polyA cassette of pD383/puro was cloned as a XhoI restriction fragment into a unique Sad site of FlipROSA $\beta$ geo, eFlip3ROSA $\beta$ geo and eFlip6ROSA $\beta$ geo to obtain of FlipROSA $\beta$ geoPuro, eFlip3ROSA $\beta$ geoPuro and eFlip6ROSA $\beta$ geoPuro, respectively (SEQ ID NOs:6-8).

C. Isolation of Puromycin Resistant Clones and Quantification of Gene Trap Events by LacZ Staining.

Generation of retroviral particles: Virus producer cells ( $4 \times 10^5$ ) (Kinsella, T. M., Nolan, G. P., Hum. Gene Ther. 7:1405-13 (1996)) were seeded onto P90 dishes and grown in DMEM (high glucose) supplemented with 10% FCS, 20 mM glutamine, 1x non-essential amino acids, and 0.1 mM  $\beta$ -mercaptoethanol. After incubating for 3 days, the cell were transfected with the retroviral constructs using Lipofectamin 2000 (Invitrogen) and the manufacturers instructions. 48 hours later, cells were overlaid with 5 ml or ES cell medium (see below) and virus particle containing supernatants were harvested after incubating for 4.5 hours. Supernatants were filtered through a 0.45  $\mu$ m Millipore filter, supplemented with 5  $\mu$ g/ml Poybrene and stored at -80° C. until use.

ES cell cultures: 129/Sv/C57BL6 F1-ES-cells were grown in DMEM (high glucose) supplemented with 15% FCS, 2 mM glutamine, 20 mM HEPES, 1 mM sodium pyruvate, 1x non-essential amino acids, 0.1 mM mercapto-ethanol, and 1500 U/ml leukemia inhibitory factor (LIF) (Chemicon).

Infection of mouse embryonic stem cells:  $1 \times 10^5$  ES cells were seeded onto gelatinized P60 Petri dishes and allowed to attach overnight. Cells were then exposed to 2 ml virus containing supernatant for 4.5 h. After adding 2 ml of fresh ES cell medium and incubating overnight, cells were put in selection for 12 days using ES cell medium containing either 0.6  $\mu$ g/ml puromycin.

X-Gal staining: After washing in PBS, cells were fixed in 3% formaldehyde and incubated overnight in lacZ staining buffer (5 mM K<sub>3</sub>[Fe(CN)<sub>6</sub>]), 5 mM K<sub>4</sub>[Fe(CN)<sub>6</sub>], 2 mM MgCl<sub>2</sub> in PBS) and 1.25 ml X-Gal solution (40 mg/ml in dimethylformamide).

D. Results:

Since retroviruses integrate mostly randomly throughout the genome, only a small fraction of the vectors will produce a gene trap event by inserting into an expressed gene. Gene trap events induce  $\gamma$ -galactosidase (LacZ) expression, which can be visualized by staining with X-Gal. Consistent with previous observations, less than 4% of all FlipROSA $\beta$ geoPuro insertions (=number of puromycin resistant clones) generated a gene trap event (=number of LacZ+ clones) (Table 1). In contrast, over 30% of the eFlip6ROSA $\beta$ geoPuro insertions generated a gene trap event, suggesting an ORE/Oct-4 mediated induction of gene expression at the insertion site (Table 1). Since activation can involve both expressed and non-expressed genes, the vectors of the invention will (i) improve the efficiency of trapping by decreasing the threshold for reporter protein detection, and (ii) increase the fraction of genes accessible to trapping by including silent genes.

TABLE 1

Induction of gene expression by enhanced gene trap vectors*						
	FlipROSA $\beta$ Geo		eFlip3ROSA $\beta$ Geo		eFlip6ROSA $\beta$ Geo	
Expt.	LacZ <sup>+</sup>	total	LacZ <sup>+</sup>	total	LacZ <sup>+</sup>	total
1	0	55	7	66	8	45
	1	46	5	23	12	35
2	2	35	2	32	9	35
	4	40	5	37	12	46

TABLE 1-continued

Expt.	Induction of gene expression by enhanced gene trap vectors*		FlipROSA $\beta$ Geo		eFlip3ROSA $\beta$ Geo		eFlip6ROSA $\beta$ Geo	
	LacZ <sup>+</sup>	total	LacZ <sup>+</sup>	total	LacZ <sup>+</sup>	total	LacZ <sup>+</sup>	total
3	2	157	6	95	42	127		
	7	103	5	138	41	97		
10	16	436	30	391	124	385		
			3.7%		7.7%		32.2%	

\*F1-ES cells were infected with retrovirus particle containing supernatants from FNXEco producer cells after transiently transfecting the gene trap plasmids. After selecting in 0.8  $\mu$ g/ml puromycin for 10 days, resistant clone were stained with X-Gal and counted. Results are from duplicate plates of 3 independent experiments are shown.

### Example 2

To determine whether enhanced gene trap vectors trap genes more efficiently, we produced several hundreds of ES cell clones with each of the conditional gene trap vectors—FlipRosa $\beta$ geo and eFlip6Rosa $\beta$ geo—(FIG. 4). After isolating the GTSTs by 5' RACE, we determined the number of genes trapped by each vector that had not been trapped before with the collection of standard gene trap vectors used by the German Gene Trap Consortium (GGTC).

#### A. Isolation of Mutant ES Cell Lines Expressing FlipRosa $\beta$ geo and eFlip6ROSA $\beta$ geo:

eFlip6ROSA $\beta$ geo gene trap virus was produced as described in Example 1B for the FlipROSA $\beta$ geoPuro class of vectors. ES cells were infected with the virus containing supernatants at an M.O.I.<0.5 as described in Example 1B. Gene trap expressing ES-cell lines were selected in 130  $\mu$ g/ml G418 (Invitrogen), manually picked, expanded, and stored frozen in liquid nitrogen until use.

#### B. Recovery of GTSTs by 5' RACE and Database Analysis:

cDNAs were prepared from the polyadenylated RNA using a RoboAmp robotic device (MWG Biotech, Ebersberg, Germany) with a processing capacity of 96 samples/day. Samples of  $2 \times 10^5$  cells were lysed in 1 ml of lysis buffer containing 100 mM Tris/HCl pH 8.0, 500 mM LiCl, 10 mM EDTA, 1% LiDS and 5 mM DTT. Polyadenylated RNA was captured from the lysates by biotin-labeled oligo-d(T)-primers according to the manufacturers instructions (Roche Diagnostics Corp., Indianapolis, Ind., USA) and placed on streptavidin-coated 96-well plates (AB Gene, Surrey, UK). After washing, solid-phase cDNA synthesis was performed in-situ using random hexamers and SuperScript II RT (Invitrogen, Karlsruhe, Germany). To remove excess primers the cDNAs were filtered through Multiscreen PCR plates (Millipore Corp. Bedford, Mass., USA). The 5' ends of the purified cDNAs were tailed with dCTPs using terminal transferase -TdT- (Invitrogen, Karlsruhe, Germany) following the manufacturer's instructions.

For PCR amplification of GTSTs, the following vector-specific primers were used: 5'-CTA CTA CTA CTA GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG-3' (SEQ ID NO:21) and 5'-GCC AGG GTT TIC CCA GTC ACG A-3' (SEQ ID NO:22); 5'-CTA CTA CTA CTA GGC CAC GCG TCG ACT AGT AC-3' (SEQ ID NO:23) and 5'-TGT AAA ACG ACG GCC AGT GTG AAG GCT GTG CGA GGC CG-3' (SEQ ID NO:24) (nested). Amplification products were directly sequenced using AB377 or ABI3700 sequencing machines (Applied Biosystems ABI, Foster City, USA).

After filtering sequences against repeats and removing all vector sequences from the GTSTs, a PHRED score was

assigned to each individual nucleotide. GTSTs qualified as informative if they were at least 50 nucleotides long and exhibited a minimum mean PHRED score of 20 (FIG. 1, supplementary information). Homology searches were performed using the publicly available sequence databases and the BlastN algorithm. Databases included GenBank, UniGene, OMIM (all at <http://www.ncbi.nlm.nih.gov>), ENSEMBL (<http://www.ensembl.org>), RIKEN (<http://www.rarf.riken.go.jp>) and GeneOntology (<http://www.geneontology.org>).

### C. Results:

GGTC's library is presently the largest public library of ES cell lines with gene trap insertion in single genes. It presently contains 13,616 GTSTs ([www.genetrap.de](http://www.genetrap.de)) corresponding to 3,349 unique genes. Table 2 shows, that 15% of the genes trapped by eFlipRosa $\beta$ geo were not present in the database, i.e. had not been trapped before, which is almost twice as much the number of novel genes trapped by standard FlipRosa $\beta$ geo. This suggests that in ES cells the genomic target accessible to enhanced gene trap vectors is significantly larger.

TABLE 2

Trapping efficiency of novel genes by enhanced gene trap vectors as estimated by 5'RACE*		
Vector	GTSTs	Novel genes (Refseq)
FlipRosa $\beta$ geo	1,766	145 (8%)
eFlip6Rosa $\beta$ geo	1,187	174 (15%)

\*GTSTs recovered from FlipRosa $\beta$ geo and eFlip6Rosa $\beta$ geo expressing ES cell clones were aligned to the GGTC-GTST database using the BlastN algorithm

This conclusion was re-enforced by determining the number of sentinel genes trapped by the two vectors that were not present in any gene trap resource, including the International Gene Trap Consortium's (IGTC) database with 27,000 GTSTs and Lexicon Genetics' Omnipbank with 200,000 GTSTs (all available at NCBI). "Sentinel" genes are fully genome annotated genes, which are presently at 7,984 (Skarnes, W. C. et al., Nat. Genet. 24:13-4 (2000)). Table 3 shows that eFlipRosa $\beta$ geo trapped sentinel genes not present in both resources about 40% more efficiently than FlipRosa $\beta$ geo.

TABLE 3

Trapping efficiency of novel sentinel genes by enhanced gene trap vectors*			
Vector	Sentinel genes	not trapped by IGTC	not trapped by Lexicon
FlipRosa $\beta$ geo	265	26 (29.8%)	7 (2.6%)
eFlipRosa $\beta$ geo	243	43 (17.7%)	11 (4.5%)

\*GTSTs recovered from FlipRosa $\beta$ geo and eFlip6Rosa $\beta$ geo expressing ES cell clones corresponding to sentinel genes were aligned to the IGTC's and Lexicon's databases using the BlastN algorithm.

### Example 3

The large scale recovery of GTSTs from mutant ES cell lines produced with conventional gene trap vectors relies on the PCR amplification of fusion transcripts using semiautomatic 5'RACE. The method, while generally robust, is dependent on transcript levels, which if too low preclude PCR amplification. In contrast to conventional gene trap expressing clones for which over 80% of RT-PCR amplifications were successful, less than 50% of the eGTV expressing clones gave meaningful 5' RACE amplification products,

suggesting that the eGTV insertions occurred into either silent or weakly expressed genes. To test this hypothesis, we subjected 51 ES cell lines that failed to generate amplification products to genomic (inverse) PCR.

#### A. Recovery of GTSTs by Inverse PCR:

Genomic DNA was isolated using the DNeasy kit of Quiagen according to the manufacturers protocol. DNA was eluted in 150  $\mu$ l. Approximately 3  $\mu$ g genomic DNA were digested in 100  $\mu$ l with 20 u NspI at 37° C. overnight. 10 Digested DNA was purified using the Qiaquick kit according to the manufacturers protocol, ligated in 300  $\mu$ l at 16° C. overnight and again purified using the Qiaquick kit.

5' inverse PCRs were carried out using the oligonucleotides I16 (5'-CGA GCC CCA GCT GGT TCT TTC-3'; SEQ ID 15 NO:25) and SR1 (5'-GCT AGC TTG CCA AAC CTA CAG GTG G-3'; SEQ ID NO:26). Nested PCR was carried out using the oligonucleotides I15 (5'-GTC TCA GAA GCCATA GAG CCC-3'; SEQ ID NO:27) and SR2 (5'-GCC AAA CCT ACA GGT GGG GTC TTT-3'; SEQ ID NO:28). 3' inverse 20 PCR was carried out using the oligonucleotides I14 (5'-ACT ATC CCG ACC GCC TTA CTG C-3'; SEQ ID NO:29) and iPCRu3 (5'-CCT CCG ATT GAC TGA GTC GCC C-3'; SEQ ID NO:30). Nested PCR was carried out using the oligonucleotides I13 (5'-TGT TTT GAC CGC TGG GAT CTG C-3'; 25 SEQ ID NO:31) and iPCRu4 (5'-TAC CCG TGT ATC CAA TAA ACC C-3'; SEQ ID NO:32).

#### B. Results:

Sequencing of the amplification products showed that of 25 eGTV insertions in annotated genes, 15 (60%) were novel 30 and not present in GGTC's database, suggesting that the novel genes were either silent or poorly transcribed prior to insertion.

TABLE 3

eFlip6Rosabgeo insertions into annotated genes identified by inverse PCR			
Clone	Chromosome	Gene	Novel*
M103A02	5	Add1	no
M103A06	7	ENSMUSG00000036862	YES
M103B03	1	Fbxo36	YES
M103B04	17	ENSMUSEST00000012809 + ENSMUSEST00000012808	no
M103B05	12	SERINE PALMITOYL-TRANSFERASE 2	no
M103B06	4	Perlecan	no
M103B07	11	POTENTIAL HELICASE WITH ZINC-FINGER DOMAIN	no
M103C01	7	ENSMUSESTG0000006748	no
M103C02	7	ENSMUSG00000007833	YES
M103C05	17	1700061G19Rik	YES
M103C07	12	SERINE PALMITOYL-TRANSFERASE 2	no
M103D01	7	ENSMUSESTT00000023443	YES
M103D03	5	Q8C4V2	YES
M103D04	5	SBB126 HOMOLOG	YES
M103E01	17	MYELIN-OLIGODENDROCYTE GLYCOPROTEIN PRECURSOR	YES
M103E02	11	TUMOR DIFFERENTIALLY EXPRESSED PROTEIN 1	YES
M103E03	19	ENSMUSESTG00000018244	YES
M103E05	11	Pled3	YES
M103E06	12	SERINE PALMITOYL-TRANSFERASE 2	no
M103E08	12	Actn1	no
M103F02	11	B230379M23Rik	no
M103F03	7	ENSMUSESTG00000019374, TYPE I INOSITOL-1,4,5-TRISPHOSPHATE 5-PHOSPHATASE	YES
M103F04	7	2410004H02Rik, Aldehydedehydrogenase	YES

17

TABLE 3-continued

eFlip6Rosabgeo insertions into annotated genes identified by inverse PCR			
Clone	Chromo-some	Gene	Novel*
M103F05	8	ENSMUSESTG00000004620	YES
M103F06	8	GH REGULATED TBC PROTEIN 1	YES

\*refers to genes not present in the GGTC database.

#### Example 4

This example describes the comparative rate of trapping of the FlipROSA $\beta$ geo and eFlip6ROSA $\beta$ geo gene trap vectors. FIG. 5 shows that at average eFlip6ROSA $\beta$ geo traps a novel gene in with every 1.4 insertions. In contrast, the non-enhanced FlipROSA $\beta$ geo requires 2.8 insertions to trap a novel gene, clearly indicating that the enhancer elements in the eFlip6ROSA $\beta$ geo gene trap vector significantly improve the efficiency of trapping.

#### Example 5

This example describes another enhanced gene trap vector of the invention containing the cytomegalovirus (CMV) immediate early enhancer which is essentially ubiquitous. This enhancer has been shown to activate gene expression in embryonic stem cells (Chung, S. et al., Stem Cells 20:139-45 (2002)) and in all mouse tissues in either orientation and up to a distance of several kbp from the promoter (Dorsch-Hasler,

K. et al., Proc. Natl. Acad. Sci. USA 82:8325-9 (1985)). The CMV enhancer has been used in combination with the chicken 6-actin promoter to drive the expression of transgenes in the mouse (Rodriguez, C. I. et al., Nat. Genet. 5:25:139-40 (2000); Zong H. et al., Cell 121:479-492 (2005); Okabe M. et al., FEBS Lett. 407:313-319 (1997)).

#### Example 6

10 This example describes another enhanced gene trap vector of the invention containing the enhancer from the mouse embryonic stem cell virus. This virus is a synthetic retrovirus derived from a mutant myeloproliferative sarcoma virus (PCMV). The enhancer element of this virus is demonstrated activity in embryonic carcinoma cells as well as embryonic stem cells<sup>21</sup>.

#### Example 7

20 This example described the use of trapped ES cell lines for making mutant mice. ES-cell derived chimeras were generated by injecting C57Bl/6 blastocysts with ES cells from the following trapped lines as obtained in Example 2: P015F03 P016F03, P023A01, P023F01, Q001D04, and Q016D06. Male chimeras were obtained with each clone and were bred to C57Bl/6 females. Litters were analyzed for germline transmission using the agouti coat color marker and tail blotting. So far, the clones P015F03 and P016F03 generated transmitted the mutation to the F1 generation. F1 mice were inter-crossed to obtain homozygous (mutant) F2 offspring for phenotype analysis.

---

#### SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 33

<210> SEQ ID NO 1
<211> LENGTH: 8586
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Retroviral vector FlipROSAbetaGeo(int)

<400> SEQUENCE: 1

ctgcagccctg aatatggccc aaacaggata tctgtggtaa gcagttcctg ccccggtca 60
gggccaagaa cagatggAAC agctgaatat gggccaaaca ggatatctgt ggtaAGCAGT
tcctgccccg gctcaggccc aagaacagat ggtccccaga tgcggccag ccctcagcag 120
tttcttagaga accatcagat gtttccaggg tgcccaagg acctgaaatg accctgtgcc 180
ttatTTGAAC taaccaatca gttcgcttct cgcttctgtt cgccgcgttc tgctccccga 240
gctcaataaa agagcccaca acccctcaact cggggcgcca gtcctccgat tgactgagtc 300
gcccggtatcc ccgtgtatcc aataaacctt ctgcagttt catccgactt gtggctcgc 360
tgttccttgg gagggtctcc tctgagtgtat tgactaccgg tcagcggggg tcttcattt 420
gggggctcgat ccgggatcg gggccctgg cccaggggacc accgaccac caccggggagg 480
caagctggcc agcaacttat ctgtgtctgt ccgattgtct agtgtctatg actgatTTTA 540
tgccgcgtcg tcggtaact agctctgtat ctggccggacc cgtgggtggaa 600
ctgacgagtt cggAACACCC ggccgcacc ctggggagacg tcccaggggac ttccggggcc 660
gtttttgtgg cccgacctga gtccaaaaat cccgatcggtt ttggacttctt tggtgcaccc 720
cccttagagg agggatatgt gggttctggta ggagacgaga acctaaaaaca gttcccgct 780
840

```

-continued

---

ccgtctgaat ttttgttgc ggttggac cgaagccgac ccgcgcgtct tgtctgtgc	900
agcatcggttc tgggtgtct ctgtctgact gtgtttgtt atttgtctga gaatttagggc	960
cagactgtta ccactccctt aagtttgacc tttaggtact ggaaagatgt cgagcggatc	1020
gctcacaacc agtcggtaga tgtcaagaag agacgttggg ttaccttctg ctctgcagaa	1080
tggccaacct ttaacgtcgg atggccgcga gacggcacct ttaaccgaga cctcatcacc	1140
caggtaaga tcaaggctt ttcacctggc ccgcacatggc acccagacca ggtccccatc	1200
atcgtgaccc gggaaacccctt ggctttgac cccctccctt gggtaaagcc ctttgcacac	1260
cctaaggcctc ccgcctccctt tcctccatcc gccccgtctc tcccccttga accttcctgt	1320
tcgacccccc ctcgatccctc cctttatcca gccccactc ctttctctagg cgccggccgg	1380
atccgttgcgt gaagtttctt ttcccaagtt cctatttttc aaatagtata ggaacttcgt	1440
taacgaagtt octattccga agtttcttatt ctctagaaag tataggaact tctcggtacg	1500
gtctcgaagc cgcgggtggcgt gtgcggggcgt gtgcgggtggcgt gtcggggggcgtactcc	1560
acctcaccctt tctggtccac gggatcccag tgggtggta ctcgaggtcg actctagagg	1620
atcgagcccc agctgggtctt ttccgtctca gaagccatag agccccccgc atccccagca	1680
tgcctgttat tggcttccca atccctcccc ttgtgttgcctt gcccccccccc accccccaga	1740
atagaatgac acctactcag acaatgcgtat gcaatttccctt cattttatata gaaaggaca	1800
gtggggagtgg cacccatccag ggtcaaggaa ggcacgggggg aggggcaaac aacagatggc	1860
tggcaacttag aaggcactgt cgaggctgtat cagcgtgtcc tagagaatttgc atccccctcag	1920
aagaactcgtt caagaaggcg atagaaggcg atgcgtgtcg aatcgggagc ggcgtatccg	1980
taaagcacga ggaagccgtc agcccatccg ccgcacatgt ctccagcaat atcacgggtt	2040
gcacaaegcta tggctgtata ggggtccgc acacccagcc ggcacacatgt gatgtatcca	2100
aaaaagccgc cattttccac catgatattc ggcaagcagg catgcgtatgt ggtcacgc	2160
agatccctcgcc cgtcgccat gcggcccttg agccctggcg acagttccgc tggcgccgc	2220
ccctgtatgtt ctccgtccat atcatccgtat tcgacaagac cggcttccat ccgagttacgt	2280
gctcgctcgat tggatgtttt cgcttgggtgg tggatgggg aggttagccgg atcaagcgta	2340
tgcageccgcgc gcatgtatgc agccatgtatgc gatactttcc cggcaggaggc aaggtggat	2400
gacaggagat cctgccccgg cacttcgcac aatagcagcc agtcccttcc cgcttcgt	2460
acaacgtcgat gcaacggcgcc cccgtcgatgg ccagccacga tagccgcgt	2520
gcctcgatgtt gcagttccat cagggcaccg gacaggtcgatgg tcttgcacaaa aagaaccggg	2580
cgccccctcgat ctgacacggcg gaacacggcg gcatcgtatgc agccgttgcgt ctgttgcgc	2640
cagtcatatgc gcaatagccat ctccacccaa gggccggag aacctgcgtg caatccatct	2700
tgttcaatgg ccgatccat attggctgc gccccggggg tccccgtaca ccagaccaac	2760
tggtaatggt agcgacccggc gtcagatgtt aattccggcg atactgtacgg gtcggcggag	2820
tgcgtccac caatccat attggaaacccg tcgatattca gccatgtgcc ttcttcgcgc	2880
tgcagcagat ggcgtatggctt ggttccatc agttgtgtt gactgtatgc gtcgtatgttgc	2940
aactggaaatgtt cggccgcgc gtcgggtggg ccataattca attcgccgtt cccgcacgc	3000
agaccgtttt cgctcgccaa gacgtacggg gtatacatgt ctgacaatgg cagatcccg	3060
cggtcaaaac aggccggcgtt aaggccggcgtt ggatagttt ctgcggccca taatccgac	3120
cagtttaccc gctctgtatc ctgcggccatc tggcgttca ggcacatccg cgccggatgc	3180

-continued

---

ggtgtatcgc tcgccacttc aacatcaacg gtaatcgcca tttgaccact accatcaatc	3240
cggttagttt tccggctgat aaataagggtt ttcccctgat gctgccacgc gtgagcggtc	3300
gtaatcagca ccgcattcagc aagtgtatct gccgtgcact gcaacaacgc tgctcgcc	3360
tggtaatggc ccgcgcctt ccagcggtcg acccaggcg tagggtcaat gcgggtcgct	3420
tcacttacgc caatgtcggt atccagcggt gcacgggtga actgatcgcg cagcggcg	3480
agcagttgtt ttttatcgcc aatccacatc tgtgaaagaa agcctgactg gcggtaaat	3540
tgccaaacgct tattacccag ctcgatgcaa aaatccattt cgctgggtgg cagatcgccc	3600
atggcgtggg acgcggcggg gagegtcaca ctgagggtt ccgcccagacg ccactgctgc	3660
caggcgctga tgtgccccggc ttctgaccat gcggtcgctg tcgggtgcac tacgctact	3720
gtgagccaga gttgccccggc gctctccggc tgcggtagtt caggcagtc aatcaactgt	3780
ttaccttgcg gagegtcaca cagaggcact tcaccgttgc ccagcggctt accatccagc	3840
gcaccatcc agtgcaggag ctcgttatcg ctatgacgga acaggtatcc gctggtact	3900
tcgatggttt gccccggataa acggaactgg aaaaactgct gctgggttt tgcttccg	3960
agcgctggat gcgccgtgcg gtcggcaag accagaccgt tcatacagaa ctggcgtatcg	4020
ttcggcgtat cgccaaaatc accggccgtaa gccgaccacg gggtgcccgtt ttcatcatat	4080
ttaatcagcg actgatccac ccagtccac acgaagccgc cctgtaaacg gggatactga	4140
cgaaacgcct gccagtatcc agcgaaaccc ccaagactgt taccatcgc gtggcgat	4200
tcgcaaaggc tcagcgccgc cgtctctcca ggttagcgaa gccatfffft gatggacat	4260
tccggcacag cccggaaaggg ctggcttca tccacgcgcg cgtacatcg gcaaataata	4320
tcgggtggccg tgggtcgcc tccggccct tcatactgca cccggcgaaa aggatcgaca	4380
gatttgcattcc agcgatacag cgcgtcgta ttgcgcgttgc ggcctgatcc attccccagc	4440
gaccagatga tcacactcggt gtgattacga tcgcgtgc ccattcgctg tacgcttcg	4500
ctcatcgccg gtagccagcg cggatcatcg gtcagacgat tcattggcac catggcg	4560
gtttcaatat tggcttcatc caccacatac agggcgtagc ggtgcacag cgtgtaccac	4620
agcggtatgt tcggataatg cgaacagcgc acggcgatcc agttgttctg cttcatcagc	4680
aggatatactt gcaccatcgat ctgtctatcc atgacctgac catgcagagg atgatgctcg	4740
tgacggatcc cgcctcgaaat cagcaacgcg ttgcgcgttca gcagcgcg accatccat	4800
atccgcaccc cgcggaaacc gacatcgacg gttctgtttt caatcagcgt gccgtcgcc	4860
gtgtcgatcc caaccaccgc acgatagaga ttgcggattt cggcgatcca cagttcgcc	4920
ttttcgacgt tcagacgtatgt tgacgcgtca tcggcataac caccacgctc atcgataatt	4980
tcaccgcgcg aaggcgccgt gccgcgtggc acctgcgtt caccctgcataaaagaaact	5040
gttaccgcgtt ggttagtcacg caactcgccg cacatctgaa cttcagccctc cagtagcgc	5100
cggctgaaat catcattaaa gcgagtgccaa acatggaaat cgctgatcc tggatcggt	5160
ttatgcagca acgagacgtc acggaaaatg ccgctcatcc gccacatatac ctgatctcc	5220
agataactgc cgtcacttca acgcagcacc atcaccgcgaa ggccgttttc tccggcgat	5280
aaaaatgcgc tcaggtcaaa ttcaagacggc aaacgactgt cctggccgtt accgaccac	5340
cgcgcgttgc accacagatg aaacgcccgg ttaacgcctt caaaaataat tcgcgtcg	5400
ccttcctgttgc cccatcgatcc atcaacatta aatgtgacgtc agtaacaacc cgtcgat	5460
tccgtggaa caaacggccgg attgaccgtt atggatagg ttacgttgcgtt gtagatggc	5520
gcatcgtaac cgtcgatctg ccagtttgcg gggacgcacgca cagttcgcc ctcaggaa	5580

-continued

---

tcgcactcca gccagcttc cggcaccgct tctggtgcgg gaaaccaggc aaagcgccat	5640
tcgccattca ggctgcgcaa ctgtggaa gggcgatcg tgccggccctc ttgcgtattta	5700
cgcgcagctgg cgaaaggggg atgtgctgca aggcgattaa gttggtaac gccagggtt	5760
tcccagtca c gacgttgtaa aacgacggga tccgcatgt cacagatcat caagcttatac	5820
gataccgtcg atccccactg gaaagaccgc gaagagtttgc tcccaaccg cgagctgtgg	5880
aaaaaaaaagg gacaggataa gtatgacatc atcaaggaaa ccctggacta ctgcgcctca	5940
cagatctgca gccccgggaa tccacttagtt ctgcctcgaa gtaggaattc gtgtcatgtc	6000
ggcgacccta cgcccccaac tgagagaact caaagggttac cccagttgg gcactatctc	6060
ccgaaaaccg cttctagcaa cgaagttcct atactatttg aagaatagga acttcggaat	6120
aggaacttca gcagatctgc atcagatacc attagacata acgatagacg atacagatct	6180
gtatcgctca tcgttatgtg atgcagatct gctgaagttc ctatacttcc tagagaatag	6240
gaacttcgga aaggaaacttc gaattctcgaa gggcccccggc tccgaccagct gtgcgcata	6300
tggcttgaat ogataaaata aaagatttttta ttttagtctcc agaaaaaggg gggaatgaaa	6360
gaccccccacct gtaggtttgg caagctgca caaccctca ctgcggggcgc cagtcctccg	6420
attgactgag tccgggggtt acccggttat ccaataaaacc ctcttgcagt tgcacccgac	6480
ttgtggcttc gctgttcctt gggagggtct cctctgagttt attgactacc cgtcagcggg	6540
ggtcttcac atgcagcatg tatcaaaatt aatttggttt ttttttttaa gtatttacat	6600
taaatggcca tagttgcatt aatgaatcg ccaacgcgcg gggagaggcg gtttgcgtat	6660
tggcgctct tccgettcct cgctcaactga ctgcgtgcgc tccgttgttc ggctgcggcg	6720
acgcggatca gctcaactaa aggccgttatcc acggatccatc acagaatcag gggataaacgc	6780
aggaaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt	6840
gctggcggtt ttccataggc tccggccccc tgacgagcat cacaatccatc gacgtcaag	6900
tcagagggtgg cgaaaccgcg caggactata aagataccag gctttccccc ctggaaagctc	6960
cctcgctgcgc tctccgttcc cgaccctgcg cgttaceggaa tacctgtccg ccttctccc	7020
ttcgggaagc gtggcgctt ctcatagctc acgctgttagg tatctcagtt cggtgttaggt	7080
cgttcgtcc aagctgggtt gtgtgcacga accccccgtt cagccgcacc gctgegcctt	7140
atccggtaac tatcgctttt agtccaaaccc ggtaagacac gacttatcgc cactggcagc	7200
agccactggt aacaggatta gcacagcgcg qtatgttagg ggtgtcacag agttttgaa	7260
gtgggtggct aactacggct acactagaag gacagtattt ggtatctgcg ctctgctgaa	7320
gccagttacc ttccggaaaaa gagttggtag ctcttgcattt ggccaaacaaa ccaccgcgtt	7380
tagcggtggt tttttgtttt gcaaggcagca gattacgcgc agaaaaaaaag gatctcaaga	7440
agatccttttgc tctccgttcc cgccgttgcg cgtcgttgcg aacgaaaact cacgttaagg	7500
gattttggtc atgagattat caaaaaggat ctccacccat atccctttgc ggccgcaaat	7560
caatctaaag tatatatgag taaaacttggt ctgcacgttta ccaatgttta atcagtgagg	7620
cacccatctc agcgatctgt ctatccgtt catccatgt tgccgtactc cccgtcgatgt	7680
agataactac gatacgggag ggcttaccat ctggcccccag tgctgcattg atacccgcgag	7740
acccacgctc accggctcca gatttacg caataaacca gccagccgga agggccgagc	7800
gcagaagtgg tcctgcaact ttatccgcctt ccatccagtc tattatgtt tgccggaaag	7860
ctagagtaag tagttcgcca gttaatagtt tgcccaacgt tggtgcattt gctgcaggca	7920

-continued

---

tctgttgtgtc acgctcgctcg tttggtatgg ctccattcag ctccggttcc caacgatcaa	7980
ggcgagttac atgatcccc atgttgtgca aaaaagggtt tagctccttc ggtcctccga	8040
tctgttgtcag aagaatgttgc gcccagtgt tatcactcat ggttatggca gcactgcata	8100
attctcttac tgtcatgcca tccgttaaat gctttctgt gactggtgag tactcaacca	8160
atgtcattctg agaaatgttgc atgcggcgcac cgagttgtc ttgcccggc tcaatacggg	8220
ataataccgc gccacatagc agaactttaa aagtgcctcat cattggaaaa cgttctcg	8280
ggcgaaaaact ctcaaggatc ttaccgtgt tgagatccag ttgcgtgtaa cccactcg	8340
cacccaaact atcttcagca tcttttactt tcaccagcg ttctgggtga gcaaaaacag	8400
gaaggcaaaa tgccgcggaaa aagggaataa gggcgacacg gaaatgttga atactcatac	8460
tcttcctttt tcaatattat tgaagcattt atcagggttta ttgtctcatg agcggataca	8520
tatttgaatg tatttagaaaa aataaacaaa taggggttcc ggcacatatt cccgaaaaag	8580
tgccac	8586

&lt;210&gt; SEQ ID NO 2

&lt;211&gt; LENGTH: 3728

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Retroviral vector pBABERfl

&lt;400&gt; SEQUENCE: 2

ctgcagcctg aatatggcc aaacaggata tctgttgtaa gcagttcctg cccggctca	60
gggccaagaa cagatggAAC agctgaatat gggccaaaca ggatatctgt ggtaaggcagt	120
tcctgccccg gtcaggggcc aagaacagat ggtccccaga tgcggtccag ccctcagcag	180
tttcttagaga accatcagat gttccaggg tgccccaagg acctgaaatg accctgtgcc	240
ttatTTGAAC taaccaatca gttcgtttct cgcttctgtt cgccgcgttc tgctccccga	300
gctcaataaa agagccaca accccctact cggggcgcca gtcctccgat tgactgagtc	360
gccccgggtac ccgtgtatcc aataaaccct cttgcagttt catccgactt gtggtctcgc	420
tgttccttgg gagggtctcc tctgtgtat tgacttcccg tcaagcgggg tctttcattt	480
gggggctcggt cccggatcg gggccatcg cccaggacc accgaccac caccggagg	540
caagctggcc agcaacttat ctgtgtctgt ccgattgtct agtgtctatg actgatTTA	600
tgcgcctgcg tcggacttag ttagctaact agctctgtat ctggcggacc cgtggtgaa	660
ctgacgagtt cggAACACCC gggcgcaacc ctgggagacg tcccaaggac ttccccggcc	720
gtttttgtgg cccgacctga gtccaaaaat cccgatcggtt ttggactt ttggcacccc	780
cccttagagg agggatatgt ggttctggta ggagacgaga acctaaacaa gttccccct	840
ccgtctgaat tttgttttc ggtttggac cgaagcccgcc cccgcgtct tgcgtgtgc	900
agcatacggtt tttgtgttct ctgtgtact gtgtttctgt atttgtctga gaattaggc	960
cagactgtta ccactccctt aagtttgacc ttaggtcaact ggaaagatgt cgagccgtc	1020
gctcacaacc agtcggtaga tgtcaagaag agacgttggg ttaccttctg ctctgcagaa	1080
tggccaaacct ttaacgtcgat atggcccgca gacggcacct ttaaccgaga cctctcatacc	1140
cagggttaaga tcaagggtttt ttcacctggc ccgcattggac acccagacca ggtccccctac	1200
atcggtgaccc gggaaaggctt ggcttttgcac cccctccctt gggtaagcc ctgtgtacac	1260
cctaaggctc cgcctccctt tcctccatcc gccccgtctc tcccccttga acctctcg	1320
tcggccccgc ctcgatccctc cctttatcca gcccctactc cttctctagg cggccggccgg	1380

-continued

---

atcccagtgt ggtggtaacgt aggaattctc gagggccccgg gtcgaccag ctgtgcgcac	1440
agtggcttga atcgataaaaa taaaagaggatt tattttagtct ccagaaaaag gggggaaatga	1500
aagaccccac ctgttaggtt ggcaagctag cacaaccctt cactcggggc gccagtctc	1560
cgattgactg agtcgccccgg gtacccgtgt atccaataaa ccctcttgca gttgcacatcg	1620
acttgtggtc tcgctgttcc ttgggagggt ctcccttgag tgattgacta cccgtcagcg	1680
gggggttttc acatgcagca tgtatcaaaa ttaatttggt tttttttctt aagtatttac	1740
ataaatggc catagttgc ttaatgaatc ggccaacgcg cggggagagg cggttgcgt	1800
attggcgct cttccgctt ctcgcactt gactcgctgc gtcggctgt tcggctgcgg	1860
cgagcggtat cagctcaactc aaaggcggta atacggttat ccacagaatc aggggataac	1920
geaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggcccg	1980
ttgctggcgt tttccatag gtcggccccc cctgacgagc atcacaaaaa tcgacgctca	2040
agtcagaggt ggccaaaccc gacaggacta taaagataacc aggcgtttcc ccctggaaagc	2100
tccctcggtc gtcctctgt tccgaccctg ccgcattaccg gataacctgtc cgcccttctc	2160
ccttcggaa gcggtggcgt ttctcatgc tcacgctgtt ggtatcttag ttgggtgttag	2220
gtcggtcgct ccaagctggg ctgtgtgcac gaaccccccgtt tcageccoga ccgtgcgc	2280
ttatccggta actatcgctc tgagtccaaac ccggtaagac acgacttatac gcaactggca	2340
gcagccactg gtaacaggat tagcagagcg aggtatgttag ggggtgtac agagttttt	2400
aagtggggc ctaactacgg ctacactaga aggacagtat ttggatctg cgctctgt	2460
aagccagttt ctttcggaaa aagagttgtt agctcttgcat cccgcaaaaca aaccaccgt	2520
ggtagccggc gttttttgtt ttgcaagcagc cagattacgc gcagaaaaaa aggtatctaa	2580
gaagatcctt tgatctttt tacggggctt gacgctcagt ggaacgaaaaa ctcacgttac	2640
gggatttgg tcatgagatt atcaaaaagg atcttcacccat agatcctttt gggccgc	2700
atcaatctaa agtatataatg agtaaacttg gtctgacagt taccatgtt taatcgtt	2760
ggcacctatc tcagegatct gtcttattcg ttcatccata gttgcctgac tccccgtcg	2820
gtagataact acgatacggg agggcttacc atctggcccc agtgcgtcaa tgataccgc	2880
agacccacgc tcacccggc cagatttac agcaataaac cagccagccg gaagggccga	2940
gcccggcgtt ggtccgtcaa ctttatccgc ctccatccag tctattaatt gttgcgggaa	3000
agctagagta agtagttgc cagttatag ttgcgcacaa gttgttgca ttgcgtcagg	3060
catcggtgtc tcacgctcgat cgtttggat ggcttcattt agctccgggtt cccaaacgtt	3120
aaggcgagttt acatgtatccc ccatgttgc caaaaaaaaagcg gttagtcctt tcggccctcc	3180
gatcggtgtc agaagtaagt tggccgcgtt gtatcactc atgggtatgg cagcaactgc	3240
taattctctt actgtcatgc catccgtaaatg atgcttttctt gtgactgggtt agtactcaac	3300
caagtccatc tgagaatagt gtatgcggcg accgagttgc ttttgcgggg cgtcaatacg	3360
ggataataacc gcgccacata gcagaactttt aaaagtgcctc atcattggaa aacgttctc	3420
ggggcgaaaa ctctcaagga tcttaccgtt gttgagatcc agttcgatgtt aacccactcg	3480
tgcacccaaac tgatcttcag catcttttac ttccaccaggc gtttctgggtt gagcaaaaaac	3540
aggaaggcaa aatgcggcaa aaaaggaaat aaggcgaca cggaaatgtt gaataactcat	3600
actcttcctt tttcaatattt attgaagcat ttatcagggtt tattgtctca tgagcggata	3660
catatttggaa tttttagaa aaaataaaaca aatagggtt ccgcgcacat ttccccggaaa	3720

-continued

agtgccac

3728

```

<210> SEQ ID NO 3
<211> LENGTH: 8740
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Retroviral gene trap vector FlipROSAbetaGeo

<400> SEQUENCE: 3

ctgcagcctg aatatgggcc aaacaggata tctgtggtaa gcagttccctg cccggcgtca      60
gggccaagaaa cagatggAAC agtgaatat gggccaaaca ggatatctgt ggttaAGCAGT      120
tcctgccccg gctcaggGCC aagaacAGAT ggtccccAGA tgccgtccAG ccctcAGcAG      180
tttcttagAGA accatcAGAT gttccAGGG tgccccAAAG acctgAAATG acccttgGCC      240
ttatTTGAAC taaccaATCA gttcgCTTC cgTTCTGTt cgcgcgCTC tgctccccGA      300
gctcaataAA agagcccACA acccCTcaCT cggggcGCCA gtcctccGAT tgactgAGTC      360
gccccgggtac ccgtgtatCC aataaacCCt cttgcAGTtG catccgactt gtggTCTCGC      420
tgttccTtgg gagggTctCC tctgagTgtAT tgactacCCG tcagcggggg tctttcATT      480
ggggggctcgT cccggatCgg gagacccCTG cccaggGACC accgaccCac cacggggagg      540
caagctggCC agcaacttAT ctgtgtCTGT ccgattGTct agtgtctatG actgatttA      600
tgcgcctcgC tcggTactAG ttAGCTAact agtctgtat ctggcggACC cgtggTggAA      660
ctgacgagTT cggAACACCC ggccgcaACC ctgggagACG tcccaggGA ttcggggGCC      720
gtttttgtgg cccgacCTGA gtccaaaaAT cccgatCgtt ttggactCtt tggtgcACCC      780
cccttagagg agggatATgt ggttctggTA ggagacgAGA acctaaaACA gttccgcCT      840
ccgtctGAAT ttttgcTTc ggTTTgggAC cgaagccgcG ccgcgcgtCT tgcTGTcGTc      900
agcatcgTTc tgcgtgtCT tgcTCTGACT gtgttCTGT atttgcTGA gaattaggc      960
cagactgtTA ccactccCTT aagtTGACC ttAGGTACT ggAAAGATGT cgagcggATC      1020
gctcacaACC agtcggTAGA tgtcaAGAAG agacgttggg ttaccttCTG ctctgcAGAA      1080
tggccaaCCT ttaacgtCGG atggccgcGA gacggcacCT ttaaccgAGA cctcatcACC      1140
caggTTAAGA tcaaggTCTT ttacacTggc ccgcATggAC acccAGACCA ggtccccTAC      1200
atcgtgacCT ggaaAGCCTT ggCTTTGAC cccCTCCCT gggtcaAGCC ctTTGtACAC      1260
cctaaggCCTC cgcctccTCT tcctccatCC gccccgtCTC tccccCTTA ACCTCTCTG      1320
tcgacccCGC ctgcgtcCTC ccttTatCCA gcccTcaCTC ctTCTCTAGG cgccggCCGG      1380
atccgttgCT gaagttcCTA ttccgaaGTT CCTATTCTC AAATAGTATA ggaacttCgt      1440
taacgaaGTT CCTATTCCGA AGTCCCTATT CTCTAGAAAG TATAGGAACt tctcgTgAcG      1500
gtctcgaAGC CGCGGTggAG cgagGAAGCG gaagAGTcta GAATAACTC GTATAGTATA      1560
cattatacGA agttatGGGT cgatGGTgAG atctggACTA gaggGTGAG ggtgtatGTT      1620
ggataactc gtatacGATA cattatacGA agttatCGGA tcccAGTGTG gtggTactcG      1680
aggtcgactc tagaggatcG agccccAGtC ggttCTTCC gtctcAGAAG ccatAGAGCC      1740
caccgcATCC ccAGCATGCC tgcttattGTC ttcccaATCC tccccCTTGc tgcTCTGCC      1800
caccccaCCC CCCAGAAATAG AATGACACCT ACTCAGACAA tgcgatGCAA tttcCTCATT      1860
ttatttagAA aggacAGTGG gagtggCACt ttccAGGGtC aaggAGGCA cgggggAGGG      1920
gcaaacaACA gatggcTggc aactAGAAGG cacAGTcGAG gctgatcAGC gagctcAGA      1980
gaattgatCC cctcagaAGA actcgtcaAG aaggcGATAG aaggcGATGC gctgcGAATC      2040

```

-continued

---

gggagcggcg ataccgtaaa gcacgaggaa gcggtcagcc cattgccgc caagctttc	2100
agcaatatca cgggttagcca acgctatgtc ctgatagcgg tccgcccacac ccagccggcc	2160
acagtcgatg aatccagaaa agcggccatt ttccaccatg atattcggca agcaggcatc	2220
gccccatgggtc acgacgagat cctcgccgtc gggcatgcgc gccttgagcc tggcgaacag	2280
tccggctggc gcgagccct gatgtcttc gtccagatca tcctgatcga caagaccggc	2340
ttccatccga gtacgtgctc gtcgatgcg atgttgcgt tggtggcga atggcaggt	2400
agccggatca agcgtatgca gcccgcgtat tgcattcagcc atgatggata ctttctggc	2460
aggagaagg ttagatgaca ggagatcctg cccggcact tcgccccata gcagccagtc	2520
ccttcccgct tcagtgacaa cgtcgagcac agtcgcgeaa ggaacgccc tgcgtggccag	2580
ccacgatagc cgccgtgcct cgtctcgag ttcatcagg gcacccggaca ggtcggtt	2640
gacaaaaaga accggggcgc cctgcgctga cagccggaaac acggccggcat cagacgagcc	2700
gatttgtctgt tgtgcggcagt catagccaa tagcctctcc acccaagcgg ccggagaacc	2760
tgcgtgoaat ccattttgtt caatggccga tcccatattt gtcgcagccc gggggatccc	2820
ctgacaccag accaacttgtt aatggtagcg accggcgctc agtcggaaatt ccgcgcatac	2880
tgcgggctc caggagtcgt cgccaccaat cccatatgg aaaccgtcga tattcagcca	2940
tgtgccttct tccgcgtgca gcagatggcg atggctgggt tccatcgat gtcgttgact	3000
gtagcggctg atgttaact ggaagtcgccc ggcactgg tgcgtggccat aattcaattc	3060
gccccgtcccg cagcgcagac cgttttcgtc cggggaaacgc tacggggat acatgtctga	3120
caatggcaga tcccagcggt caaaacaggg ggcagtaagg eggtcggtt agtttcttgc	3180
cggccctaatt ccgagccagt ttacccgtc tgctacctgc gccagctggc agttcaggcc	3240
aatcccgccc ggtatgcggtg tategcgtc cacttcaaca tcaacggtaa tcgcatttgc	3300
accactacca tcaatccggt aggtttccg gctgataat aaggtttcc cctgtatgt	3360
ccacgcgtga gccccgtcaa tcagcaccgc atcagcaagt gatatgtccg tgcactgca	3420
caacgcgtct tcggcgttgtt aatggccgc cgccttcgcg cgttcgaccc aggccgtt	3480
gtcaatgcgg gtcgtttcac ttacgccaat gtcgttatcc agcgggtcaca gggtaactg	3540
atcgccgcagc ggcgtcagca gttttttt atcgccaatc cacatgtgaa aagaaagcc	3600
tgactggcggtt taaaattgcc aacgcttatt acccagctcg atgcggaaat ccatttcgt	3660
gtgtggcaga tgcggatgg cgtggggacgc ggcggggagc gtcacactga ggtttccgc	3720
cagacgcac tgcgtccagg cgctgtatgtt cccggcttgc gaccatgcgg tcgcgttcgg	3780
ttgcactacg cgtactgtga gcccagatgtt cccggcgctc tccggctgcg gtatgttgc	3840
cagttcaatc aactgtttac ctgtggcgc gacatccaga ggcacttcac cgcttgcag	3900
ccgccttacca tccagcgcac ccattccatgtc caggagctcg ttatcgctat gacggaaacag	3960
gtatcgctg gtcacttcga tggtttgcgg ggtataaacgg aactggaaaa actgtgtcg	4020
gtgttttgcg tccgtcagcg ctggatgcgg cgtgcggctcg gcaaaagacca gaccgttcat	4080
acagaactgg cgtatcgatcgc aaaatccaccc cgtatgcgc accacgggtt	4140
ccgcgtttca tcatatccaa tcagcgtactg atccacccag tcccaagacga agccgcctg	4200
taaacggggta tactgacgaa acgcctgcac gtatgttgcg aaaccgcac gactgttacc	4260
catcgctgg gctgtattcgc aaaggatcg cggccgcgtc tcccgatgtt gcaacggcc	4320
tttttgcgtt gaccatttcg gcacagccgg gaagggtcg tccatccca cgcgcgcgtt	4380

-continued

-continued

attaatgaat cgggcaacgc gccccggagag gcggttgcg tattggcgc tctccgctt  
cctcgctcac tgactcgctg cgctcggtcg ttccggctcg gcgagcggta tcagctca  
caaaggcggt aatacggtt tccacagaat caggggataa cgcaaggaaag aacatgttag  
caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgtggcg ttttcata  
ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcaagg tggcgaaacc  
cgacaggact ataaagatac caggcggttc cccctggaaat ctccctcgat cgcttcctg  
ttccgaccct ggcgttacc ggatacctgt ccgccttct cccttcggga agcgtggcg  
tttctcatag ctcacgctgt aggtatctca ttccgggtta ggtcggtcg tccaagctgg  
gctgtgtgca cgaacccccc gttcagcccg accgctgatc cttatccgt aactatcg  
tttagtccaa cccggtaaga cacgacttat cgccactggc agcaggact ggtaacagga  
tttagcagagc gaggtatgtt ggcgggtcta cagagttttt gaagtggtgg cctactac  
gctacactag aaggacagta tttgttatctt ggcgttgcgatc gaagccagtt accttcggaa  
aaagagttgg tagcttttgc tccggcaaac aaaccaccgc tgtagcggt ggtttttt  
tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct ttgatcttt  
ctacggggtc tgacgctcag ttggAACggaa actcacgtta agggattttg gtcatgagat  
tatcaaaaag gatcttaccat tagatctttt tgccggccca aatcaatcta aagtatata  
gagtaaactt ggtctgacag ttaccaatgc ttaatcagtg aggacccat ctcagcgatc  
tgtctatttc gttcatccat agttgcgtcg ctcccgatc tgtagataac tacgatc  
gagggtttac catctggccc cagtgctgca atgataccgc gagaccacgc ctcacccgg  
ccagatttat cagcaataaa ccaggccagcc ggaaggccgc agcgcagaag tggctctgca  
actttatccg cctccatcca gtctattaaat ttggccggg aagctagagt aagtagttcg  
ccaggttataa gtttgcgcaa cgttgttgc attgctgacg gcategtggt gtcacgctcg  
tgcgtttggta tggcttcatt cagtcgggtt tcccaacat caaggcgagt tacatgatcc  
ccatgttgtt gcaaaaaagc ggttagctcc ttccggctcg cgategttgtt cagaagtaag  
ttggccgcag tggatcact catggttatg gcagcaactgc ataattctct tactgtcatg  
ccatccgtaa gatgctttc tggactggt gagtactcaa ccaagtcatt ctgagaatag  
tgtatgcgc gaccgagttt ctcttgcggc ggcgtcaatac gggataatac cgcgcacat  
agcagaactt taaaagtgtt catcattggaa aacggttctt cggggcgaaa actctcaagg  
atcttacccgc tggatcact cagttcgatc taacccactc gtgcacccaa ctgatcttca  
gcacatcttta ctttccacccag cgtttctggg tgagcaaaaa caggaaggca aaatgcgc  
aaaaaggggaa taagggcgac acggaaatgt tgaataactca tactcttcct tttcaatata  
tattgaagca ttatcagggtt tattgtctc atgagcggat acatatttga atgtatattag  
aaaaataaac aaatagggggt tccgcgcaca tttccccgaa

```
<210> SEQ ID NO 4
<211> LENGTH: 8878
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Retroviral gene trap vector eFlip3ROSAbetaGeo
```

<400> SEQUENCE: 4

ctgcagcctg aatatgggcc aaacaggata tctgtggtaa gcagttccctg ccccggtca 60

-continued

---

ggccaagaa cagatggAAC agctgaatat	ggccaaACA ggatatctgt	ggtaaggcgt	120
tccTccccG gtcaggGCC aagaacAGAT	ggTccccAGA tgccgtCCAG	ccctcAGCAG	180
tttctAGAGA accatcAGAT	gtttccAGGG tgccccAGG	acctgAAATG accctGTGCC	240
ttatTTGAAC taaccaATCA	gttgcTTCT cgTTCTGT	cgCGCGCTC tgccccGA	300
gtcAATAAA agageCCACa	accCCTCACT	cggggcGCcA gtcCTCCGAT	360
gcccGGGTAC ccgtGTATCC	aataaACCT	cttgCAGTTG catCGACTT	420
tgttcTTGG gagggTCtCC	tctgAGTGT	tgactACCCG tcAGCggggG	480
gggggCtCGT cCGGgATCGG	gagACCCG	cccAGGGACC accGACCCAC	540
caagCTGGCC agcaACTTAT	ctgtGTCTGT	ccgATTGTCT agtGTCTATG	600
tgcgcCTGCG tcggTACTAG	ttagCTAAct	agtcTCTGTat ctggcGGACC	660
ctgacgAGTT cggAACACCC	ggccgcaACC	ctgggAGACG tcccAGGGAC	720
gtttttGTGG cccgACCTGA	gtccAAAAT	cccgATEGt	780
cccttagAGG aggATATGT	ggttCTGGTA	ggagACGAGA acctaAAACA	840
ccgtCTGAAT ttttGTTTC	ggtttGGGAC	cgaAGCCGCG ccgcgcGTCT	900
agcateGTTC tGTGTTGTCT	ctgtCTGACT	gtgtttCTGT atttGTCTGA	960
cagactGTtA ccactCCCTT	aagtTGACC	ttaggtCACT ggAAAGATGT	1020
gctcacaACC agtCGGTAGA	tgtcaAGAAG	agacGTTGGG ttacCTCTG	1080
tggccAAACt ttaACGTcGG	atggcCGGA	gacggCACCT ttaAccGAGA	1140
caggTTAAGA tcaAGGTCTT	ttcacCTGc	ccgcatGGAC acccAGACCA	1200
atcgtGACCT gggAACCTT	ggctttGAC	ccccCTCCt ggtcaAGCC	1260
cctaAGCCTC CGCCTCCTCT	tcctCCatCC	gccccGTCt	1320
tcgACCCCGC ctgcAtCCTC	ccttatCCA	gcccTCactC	1380
atccGTTGCT gaAGTTCTA	ttccGAAGTT	cTTattCTC aaATAGTATA	1440
taacGAAGTT CCTATTCCGA	agttCCTATT	ctctAGAAAG tatAGGAACt	1500
gtctcGAAGC CGCGGTGGAG	cgAGGAAGCG	gaAGAGTCTA	1560
cattatacGA agttatGGGT	cgatGGTAG	atctGGACTA	1620
ggataACTTC gtATAGCATA	cattatacGA	agttatCggA	1680
aggtcGACTC tagAGGATCG	agccccAGCT	ggttCTTCC	1740
caccGcatCC CcAGCATGCC	tgcttattGTC	ttccCAATCC	1800
cacCCCACCC CCCAGAAATAG	aatGACACCT	actcAGACAA	1860
ttatTTAGGAA aggACAGTGG	gagtGGCACC	ttccAGGGTC	1920
gcaaACAACA gatGGCTGGC	aactAGAAAG	cacAGTCGAG	1980
gaattGATCC OCTCAGAAAGA	actcGTCAAG	aaggcGATGC	2040
gggAGCGCGC ataccGtAA	gcaGAGGAA	gCGGTcAGC	2100
agcaATATCA CGGGTAGCCA	acgCTATGTC	ctgAtAGCG	2160
acAGTCGATG AATCCAGAAA	agcGGCCATT	ttccACCATG	2220
gcccAtGGTC acgAcGAGAT	cctcGCCGTC	ggcAtGCGC	2280
tccGGGTGc GCGAGCCCT	gatGCTCTTC	gtccAGATCA	2340
ttccCATCCGA gtAcGtGCTC	gctcGATGEG	atgtttCGCT	2400
agccGGATCA agcGTATGCA	gccGCCGCA	tgcAtCAGCC	2460

-continued

---

aggagcaagg tgagatgaca ggagatcctg cccggcaact tcgccaata gcagccagtc	2520
ccttcccgt tcaagtacaa cgctcgagcac agctgcgc当地 ggaacgccc当地 tcgtggccag	2580
ccacgatagc cgcgctgc当地 cgtctgc当地 ttcatcagg gcacccgaca ggtcggtctt	2640
gacaaaaaga accggggcgcc cctgcgctga cagccggaaac acggcgccat cagacgagcc	2700
gattgtctgt tgcgtccagg catagccaa tagcctctcc acccaagcgg ccggagaacc	2760
tgcgtcaat ccatctgtt caatggccga tccatattg gctgcagccc gggggatccc	2820
ctgacaccag accaacttgtt aatggtagcg accggcgctc agctggaaatt ccggcgatac	2880
tgcgtggctc caggagtcgt cgccaccaat cccatattg aaaccgtcga tattcagcca	2940
tgtgccttct tccgtgtca gcagatggcg atggctggtt tccatcagtt gctgttgact	3000
gtagcggctg atgttaact ggaagtgc当地 gegccactgg tgcgtggccat aattcaattc	3060
gwgctcccg cagcgcagac cggttcgtc cgggaagacg tacggggat acatgtctga	3120
caatggcaga tcccgccgt caaaacagggc ggcaatggg cggcgatgggat agtttcttg	3180
cgccccataat ccgagccagt ttacccgctc tgctacctgc ggcagctggc agttcaggcc	3240
aatcccgccg ggtgcgttgcg tategc当地 cacttcaaca tcaacggtaa tgcgcatttg	3300
accactacca tcaatccggg aggtttccg gctgataat aagggtttcc octgatgctg	3360
ccacgctgta gcccgtctaa tcagcacccg atcagcaagt gtatctgc当地 tgactgc当地	3420
caacgctgct tccgtgttgcg aatggcccg cgcctccag cgatcgaccc aggccgttgg	3480
gtcaatgc当地 gtcgtttcac ttacgc当地 gtgc当地 agcgggtc当地 gggtaactg	3540
atcgc当地 cgc当地 ggtgc当地 gttttttt atcgc当地 acatctgtg aaagaaagcc	3600
tgactggccg ttaaaattgcc aacgcttatt acccagctcg atgaaaaat ccattcgct	3660
ggtggtcaga tgc当地 ggtggatgg cgtgggacgc ggccggggacgc gtc当地 actgta	3720
cagacgc当地 tgctgc当地 cgc当地 gttggatgg cccggcttcc gaccatgc当地 tgc当地	3780
ttgc当地 ctgc当地 gtc当地 gtc当地 cccggcgctc tccggctgc当地 gtagttcagg	3840
cagttcaatc aactgtttac cttgtggacg gacatccaga ggc当地 ctc当地 cgcttgc当地	3900
cggttacca tccagcgcc local ccatccagtg caggagctcg ttatcgctat gacggaaacag	3960
gtatcgctg tcaacttc当地 tggttgc当地 ggataaacgg aactggaaaa actgtgtcg	4020
gtgtttcgct tccgtc当地 ctggatgc当地 cgtgc当地 gcaaaagacca gaccgttcat	4080
acagaactgg cgatcgctcg cgctatcgcc aaaatcaccc cgc当地 agcccggtt	4140
gccc当地 tcaatattaa tcagcgactg atccacccag tcccaagacga agccgc当地	4200
taaacggggta tactgacgaa acgc当地 gca gtatcgatcg aaaccgc当地 gactgttacc	4260
catcgctgg gc当地 tccatccatc当地 aaaggatcg cggccgctc tccatccatc当地 gcaaaagcc	4320
tttttgc当地 gaccatccatc当地 gc当地 acgacccgg gaagggtcg tccatccatc当地 cgccgctc当地	4380
catcgcccaa ataataatcg tggccgttgcg tgc当地 cccctcat actgc当地	4440
gc当地 ggaaaggta tc当地 acagatt tgatccagcg atacagcgccg tgc当地 ggcc	4500
tgattcatc当地 occagcgacc agatgtatcg actcggtcg ttc当地 catcgatcg gctgc当地	4560
tgc当地 gttacg cgttc当地 ctgc当地 ggatcg ccagcgccg tccatcgatcg gacgattcat	4620
tggcaccatc当地 cgc当地 ggggtttt caatattggc ttc当地 atccacc acatacaggc cgtgc当地	4680
gc当地 acagcgctg taccacagcg gatgggttccg ataatgc当地 cagcgccacgg cgtt当地 aagtt	4740
gttctgctt当地 atcagcgacg ttc当地 ctgc当地 catcgatcg tccatccatc当地 cctgaccatc当地	4800

-continued

---

cagaggatga tgctcggtac ggttaacgccc tcgaatcagc aacggcttgc cgttcagcag	4860
cagcagacca tttcaatcc gcaccccgccg gaaaccgaca tcggcaggctt ctgcttcaat	4920
cagcgtgccg tcggcggtgt gcagttcaac caccgcacga tagagattcg ggatttcggc	4980
gctccacagt ttccgggtttt cgacggttcag acgttagtgc acgcgtatcg cataaccacc	5040
acgctcatcg ataatttcac cgccgaaaagg cgccgtgcgc ctggcgactt gcgttccacc	5100
ctgccataaa gaaactgtta cccgttagta gtcacgcacca tcggcgacaca tctgaacttc	5160
agcctccagt acagegcggc tgaaatcatc attaaagcga gtggcaacat ggaaatcgct	5220
gatttgtta gtcgggttat gcagcaacga gacgtcacgg aaaatgcgc tcateccgcca	5280
catatcctga tcttcagat aactgccgtc actccaacgc agcaccatca ccgcgaggcg	5340
gttttctccg gcgcgtaaaa atgcgctcag gtcaaattca gacggcaaac gactgtctg	5400
gccgttaaccg acccagcgcc cggtgcacca cagatgaaac gccgagttaa cgccatcaaa	5460
aataattcgc gtctggcctt cctgttagcca gcttcatca acattaaatg tgagcgagta	5520
acaacccgtc gatttctccg tgggaacaaa cggcggatgg accgtaatgg gataggttac	5580
gttgggttag atggggcgcat cgtaaccgtg catctgcaccc tttgagggga cgacgacagt	5640
atcggcctca ggaagatcgc actccagcca gcttccggc accgcttctg gtgcccggaaa	5700
ccaggcaaaag cgccattcgc cattcaggct gcgcactgt tgggaaggcc gatcggtgcg	5760
ggccttccg ctattacgccc agctggcgaa agggggatgt gctgcaagcc gattaagtt	5820
ggtaacgcgc gggtttcccc agtacgcacg ttgtaaaacg acgggatccg ccatgtcaca	5880
gatcatcaag cttatcgata ccgtcgatcc ccactggaaa gaccgcgaag agtttgcct	5940
caaccgcgag ctgtggaaaa aaaaggaca ggataagtat gacatcatca aggaaaccct	6000
ggactactgc gccctcaga tctgcagccc gggggatcca ctagttctag cctcgagtag	6060
gaattcgata acttcgtata atgtatacta tacgaagtta tgggtcgatg gtgatgctt	6120
gcaattcggg tcgatggta agcttggata acttcgtata atgtatgcta tacgaagtta	6180
tcaattcgcc ctatagttag tcgtatttcc ccgaaaaccg cttctagccaa cgaagttcct	6240
atactatgg aagaatagga acttcggaaat aggaacttca gcagatctgc ttttccattt	6300
gcatttcaaa tgtgatttttta acttacatta tacaaggctg aaaggctgt gcaggatctg	6360
ctttccatt tgcatttcaa atgtgattttt aacttacattt atacaaagct gaaaggctcg	6420
tgcaggatct gctttccat ttgcatttca aatgtgattt taacttacat tataacaagc	6480
tgaaaggctca gtgcaggatc tgctgaagtt cctataactt ctagagaata ggaacttcgg	6540
aataggaact tcgaatttcc gagggcccg gctcgaccag ctgtgcgcatt agtggcttgc	6600
atcgataaaa taaaagattt tatttagtct ccagaaaaag gggggatga aagacccac	6660
ctgttaggttt ggcaagctag cacaaccct cactcgcccc gccagtcctc cgattgactg	6720
agtcgcggcg gtaccctgtt atccaataaa ccctcttgc gttgcattccg acttgcgttgc	6780
tcgcgtttcc ttgggggggt ctccctctgat tgattgacta cccgtcagcg ggggtcttc	6840
acatgcgcacca tgcattcaaaa ttaatttgcgtt ttttttctt aagtatttac attaaatggc	6900
catagttgcata taaatgcaccc ggccaaacgcg cggggagagg cgggttgcgtt attggcgct	6960
cttccgccttc ctgcgtact gactcgctgc gctcggtcg tccggctgcgg cgagcggtat	7020
cagctcactc aaaggccgtaa atacggttat ccacagaatc aggggataac gcagggaaaga	7080
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaggccgcg ttgcgtggcgt	7140
ttttccatag gctccggcccc cctgacgcgcg atcacaaaaaa tcgacgcgtca agtcagaggt	7200

-continued

---

ggcgaaaccc gacaggacta taaagatacc aggcgttcc ccctggaaac tccctcgta 7260  
 getctcttgt tccgaccctg ccgcttacgg gataacctgtc cgcccttc ccttcggaa 7320  
 gctggcgct ttctcatagc tcacgctgtt ggtatctca gtcgggttag gtcgttcgt 7380  
 ccaagctggg ctgtgtcac gaacccccc ttcagccgaa ccgctgcgcc ttatccgta 7440  
 actatcgct tgagtccaac ccggtaagac acgacttatac gccactggca gcagccactg 7500  
 gtaacaggat tagcagagcg aggtatgttag gccgtgtac agagttcttg aagtgggtgc 7560  
 ctaactacgg ctacactaga aggacagttat ttggtatctg cgctctgtc aagccagtt 7620  
 cttccggaaa aagagttgt agctttgtat ccggcaaaaca aaccaccgtt ggtagcggtg 7680  
 gttttttgtt ttgcaagcag cagattacgc gcagaaaaaa aggtatctcaa gaagatcctt 7740  
 tgcgttttc tacggggct gacgctcagt ggaacgaaaa ctcacgttaa gggattttgg 7800  
 tcatgagatt atcaaaaagg atcttcaccc agatcctttt gggccgc当地 atcaatctaa 7860  
 agtatatatatg agtaaaacttg gtctgacagt taccatgtt taatcgtga ggcacccatc 7920  
 tcacgcatct gtcttattcg ttcatccata gttgcctgac tccccgttgt gtatgataact 7980  
 acgatacggg aggggttacc atctggcccc agtgctgc当地 tgataccgc当地 agacccacgc 8040  
 tcacccggctc cagatttatac agcaataaac cagccagccg gaagggccgaa gcgcagaagt 8100  
 ggtccctgcaaa ctttatccgc ctccatccag tctattaattt gttgcceggaa agttagagta 8160  
 agtagttcgc cagttaatag tttgc当地 gttgc当地 ttc当地 ctc当地 catcgtggtg 8220  
 tcacgctgtt cgttttgtat ggcttcattc agtccgggtt cccaaacgatc aagggaggtt 8280  
 acatgatccc ccatgttgc当地 caaaaagcg gttagcttcc tccgttcc gatcgttgc当地 8340  
 agaagtaagt tggccgc当地 gttatcactc atgggttatgg cagcaactgca taattcttt 8400  
 actgtcatgc catccgtaaat atgttttctt gtgactgggt agtactcaac caagtcattc 8460  
 tgagaatagt gtatgc当地 accgagttgc tcttgc当地 cgtcaatacg ggataatacc 8520  
 ggc当地 cacaatactt caaaaatgtc atcattggaa aacgttctc gggggaaaa 8580  
 ctctcaagga tcttaccgc当地 gttgagatcc agttcgatgtt aaccactcg tgcaacc当地 8640  
 tgatcttc当地 catctttac tttcaccgc当地 gtttctgggt gagcaaaaac aggaaggca 8700  
 aatgc当地 ccaaaaatggaaat aaggccgaca cggaaatgtt gaataactcat actcttc当地 8760  
 ttcaatattt attgaagcat ttatcagggtt tattgtctca tgagccgata catattgaa 8820  
 tgatatttaga aaaataaca aataggggttt ccgc当地 ccat ttccccgaaa agtgc当地 8878

<210> SEQ ID NO 5  
 <211> LENGTH: 9091  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Retroviral gene trap vector eFlip6ROSAbetaGeo

<400> SEQUENCE: 5

ctgcagcctg aatatggcc	aaacaggata tctgtggtaa	gcagttccctg ccccggtca	60			
ggccaagaa	caatggaaac	agtgaaat	ggccaaaca	ggatatctgtt	ggtaaggcagt	120
tccctggcc	ccctggcc	aagaacagat	ggtcccccaga	tgcggccag	ccctcagcag	180
tttcttagaga	accatcagat	gttccagg	tgc当地	ccctgaaatg	accctgtgcc	240
ttatggaaac	taaccaatca	gttc当地	cgcttctgtt	cgccgc当地	tgc当地	300
gtcaataaaa	agagccacaca	accctcact	cggccgc当地	gtc当地	cccgat	360

-continued

---

gcccggtac ccgtgtatcc aataaaccct cttgcagttt catccgactt gtggtctcg	420
tgttccttgg gaggggtctcc tctgagtgtat tgactacccg tcagccccggg tctttcattt	480
ggggggtcggt cgccggatcg ggagacccctg cccaggacc accgacccac caccggagg	540
caagctggcc agcaacttat ctgtgtctgt ccgatgtct agtgtctatg actgatTTA	600
tgcgcctcgcc tcggtaact agctctgtat ctggcgacc cgtggtgaa	660
ctgacgagggtt cggAACACCC ggccgcAACCC ctggggagac tcccaggaccc ttccccggcc	720
gtttttgtgg cccgacctga gtccaaaaat cccgategtt ttggacttt tggtgcaccc	780
cccttagagg agggatatgt gggtctggta ggagacgaga acctaaaaca gttcccgct	840
ccgtctgaat ttttgcTTT ggTTTGGGAC cgaagcccg CGCGCGGTCT TGTCTGCTGC	900
agcatcggttctc tgggttgcact gtgtttctgt atttgtctga gaattaggc	960
cagactgtta ccactccctt aagtttgacc ttaggtcaact ggaaagatgt cgagcggatc	1020
gctcacaacc agtcggtaga tgtcaagaag agacgttggg ttaccttcgt ctctgcagaa	1080
tggccaacct ttaacgtcggtt atggccgcga gacggcacct ttaaccgaga cctcatcacc	1140
cagggttaaga tcaagggtt ttcacctggc ccgcacggac acccagacca ggtccccctac	1200
atcgtgacccctt gggaaaggcctt ggcttttgcac ccccttcctt gggtaagcc ctttgtacac	1260
cctaaggcctc cgcctcccttc tcctccatcc gccccgtctc tcccccttgc acctcctcg	1320
tgcaccccgcc ttcgatccctc cctttatcca gcccctcaactc cttctctagg cgccggccgg	1380
atccgttgcgtt gaagttccata ttccgaagtt cctatttcata aaatagtata ggaacttcgt	1440
taacgaagtt cctattccga agttcctatt ctctagaaag tataggaact tctcgtgacg	1500
gtctcgaagc cgcgggtggag cgaggaagcg gaagagtcta gaataacttc gtatagtata	1560
cattatacga agttatgggt cgtatggtag atctggacta gagggtcgat ggtgtatgtt	1620
ggataacttc gtatagcata cattatacga agttatcgga tcccagtgtg gtggtaactcg	1680
aggtcgcactc tagaggatcg agccccagct ggttcttcc tcgtcagaag ccatagagcc	1740
caccgcattcc ccacgcattcc tgctattgtc ttcccaatcc tcccccttgc tgcctgccc	1800
cacccccc cccagaatag aatgcacaccc actcagaccaa tgcgatgcaaa ttccctcatt	1860
ttatttagaa aggacagtgg gagttggcacc ttccagggtc aaggaaggca cgggggaggg	1920
gcaaaacaaca gatggctggc aactagaagg cacagtcgag gctgtatcgc gagctctaga	1980
gaattgtatcc cctcagaaga actcgtcaag aaggcgatag aaggcgatgc gctgcgaatc	2040
ggggagccgcg ataccgtaaa gcacgaggaa gcggtcagcc cattcgccgc caagctttc	2100
agcaatatca cgggttagcca acgctatgtc ctgatagcgcc tccgcccaccc ccacccggcc	2160
acagtcgcgtt aatccagaaa agcggccattt ttccaccatg atattcggca agcaggcattc	2220
gcacatgggtc acgacgagat cctcgccgtc gggcatgcgc gcttgagcc tggcgaacag	2280
tgcgggtggc gcgagccctt gatgtcttc tgcgtatgcata tccgtatgcata caagaccggc	2340
ttcccatccga gtacgtgtc gtcgtatgcgtt atgtttcgct tgggtggtcga atgggcagg	2400
agccggatca acgctatgtca gcccggcat tgcgtatgcata atgtatggata ctttctggc	2460
aggagcaagg tgagatgaca ggagatccgt ccccgccact tcggccaaata gcagccagtc	2520
ccttcccgct tcagtgcacaa cgtcgagcac agctgcgcacaa ggaacgcccgc tcgtggccag	2580
ccacgatgc cgcgtgcct cgttcgtcgat ttccatcagg gacccggacca ggtcggtt	2640
gacaaaaaga accggggcgcc cctgcgtca cagccggaaac acggccggcat cagagcagcc	2700
gattgtctgt tgcgtccaggat catagccgaa tagcccttcacc acccaagccgg ccggagaacc	2760

-continued

---

tgcgtgcaat	ccatcttgg	aatggccga	tcccatattg	gctgcagccc	ggggatccc	2820
ctgacaccag	accaactgg	aatggtagcg	accggcgctc	agctggaa	ccggcgatac	2880
tgacgggctc	caggagtcgt	cgccaccaat	ccccatatgg	aaaccgtcga	tattcagcca	2940
tgtgccttct	tccgctgtca	gcagatggcg	atggctgggtt	tccatcgtt	gctgttgact	3000
gtagcggctg	atgttgaact	ggaagtcgc	gcccactgg	tgtggccat	aattcaattc	3060
gcccgtcccg	cagcgcagac	cgttttcgt	cggaaagacg	tacggggtat	acatgtctga	3120
caatggcaga	tcccagcggt	caaaaacaggc	ggcagtaagg	cggtcggtat	agtttcttg	3180
cggcccta	atccgagccgt	ttacccgtc	tgctacccgc	gccagctggc	agttcaggcc	3240
aatcccgcc	ggatgcgggt	tatcgctcgc	cacttcaaca	tcaacggtaa	tcgccatttg	3300
accactacca	tcaatccgg	aggtttccg	gctgataat	aaaggtttcc	cctgtatgt	3360
ccacgcgtga	gcccgtgtaa	tcagcaccgc	atcagcaagt	gtatctgcgc	tgcaactgcaa	3420
caacgcgtct	tccgcgttgt	aatggccgc	cgccttccg	cgttcgaccc	aggcgtagg	3480
gtcaatgcgg	gtcgcttcac	ttacgccaat	gtcggtatcc	agcgggtcac	gggtgaactg	3540
atcgccgc	ggcgtcagca	gttgggggtt	atcgccaatc	cacatctgt	aaagaaagcc	3600
tgactggcgg	ttaaattgcc	aacgcttatt	acccagctcg	atgcaaaaat	ccatctcgct	3660
gggtggcaga	tgccggatgg	cgtgggacgc	ggcgccccgc	gtcacactga	ggtttccgc	3720
cagacgcac	tgctgccagg	cgctgtatgt	ccggcttct	gaccatgcgg	tcgcgttcgg	3780
ttgcactacg	cgtactgtga	gccagagttg	ccggcgctc	tccggctgcg	gtagttcagg	3840
cagtcaatc	aactgtttac	cttggggcgc	gacatccaga	ggcacttcac	cgcttgcag	3900
cggcttacca	tccagcgcca	ccatccagtg	caggagctcg	ttatcgctat	gacggaaacag	3960
gtatcgctg	gtcaattcga	tggtttgc	ggataaaacgg	aactggaaaa	actgtcgct	4020
gtgtttgt	tccgtcagcg	ctggatgcgg	cgtgcggctcg	gcaaaagacca	gaccgttcat	4080
acagaactgg	cgtatgtcg	gcgtatcgcc	aaaatcaccg	ccgtaaagccg	accacgggtt	4140
gccgtttca	tcatatttaa	tcagcactg	atccacccag	tccagacga	agccgcctg	4200
taaacgggaa	tactgacgaa	acgcctgc	gtatgtcg	aaaccgc	gactgttacc	4260
catcgegtgg	gcgtattcgc	aaaggatcg	cgggcgegtc	tctccaggta	gcgaaagcc	4320
tttttgcgt	gaccatttcg	gcacagccgg	gaagggtcg	tcttcatcca	cgcgcgct	4380
catcgccaa	ataatatcg	tggccgttgt	gtcggtccg	ccgccttcat	actgcacccg	4440
gcgggaagga	tcgacagatt	tgcgtcccg	atacagcg	tcgtgattag	cgccgtggcc	4500
tgattcatc	cccagcgacc	agatgtatc	actcggtga	ttacgatcgc	gctgcacccat	4560
tcgcgttacg	cgttcgctca	tcgcggtag	ccagcgcgg	tcatcggtca	gacgattcat	4620
tggcaccatc	ccgtgggtt	caatattggc	ttcatccacc	acatacaggc	cgtacgcgtc	4680
gcacagcg	taccacagcg	gatgggtcg	ataatgcgaa	cagcgcacgg	cgttaaagtt	4740
gttctgttcc	atcagcagga	tatccgtcac	catcgctcg	tcatccatga	cctgaccat	4800
cagaggatga	tgctcggtac	ggtaacgc	tcgaatcgc	aacggcttc	cgttcagcag	4860
cagcagacca	tttcaatcc	gcacccgc	gaaaccgaca	tcgcaggctt	ctgctcaat	4920
cagcgtccgc	tcggcggtgt	gcagttcaac	caccgcacg	tagagttcg	ggattccgc	4980
gtctccacagt	tccgggtttt	cgacgttca	acgtatgt	acgcgtatcg	cataaccacc	5040
acgctcatcg	ataatttcac	cgccgaaagg	cgcgggtcc	ctggcgac	gcgtttcacc	5100

-continued

ctggccataaa gaaactgtta cccgttagta gtcacgcaac tcgcccaca tctgaacttc 5160  
agcctccagt acagegcggc tgaaaatcatc attaaagcgta gtggcaacat ggaaatcgct 5220  
gattttgtta gtcgggttat gcagcaacgaa gacgtcacgg aaaatgccgc tcatccgcca 5280  
catatccctga tcttccagat aactgccgtc actccaaacgc agcaccatca ccgcgaggcg 5340  
gttttctccg gcgcgtaaaa atgcgctcag gtcaaattca gacggcaaac gactgtcctg 5400  
gccgttaaccg acccagcgcc cggtgcacca cagatgaaac gccgaggtaa cgccatcaaa 5460  
aataattcgc gtctggccctt cctgttagcca gctttcatca acattaaatg tgagcgagta 5520  
acaacccgtc ggatttctccg tgggaacaaa cggcggattg accgtaatgg gataggttac 5580  
gttgggttag atgggcccgt catctgccag tttgagggga cgacgacagt 5640  
atccggccctca ggaagatcgc actccagcca gctttccggc accgcttctg gtgcggaaa 5700  
ccaggcaaaag cgccattcgc cattcagget ggcgaactgt tgggaaggcg gatcggtcg 5760  
ggccttctcg ctattacgccc agctggcgaa agggggatgt gtcgcaaggc gattaagttg 5820  
ggtaacgcca gggtttcccc agtcacgacg ttgtaaaacg acgggatccg ccatgtcaca 5880  
gatcatcaag cttatcgata ccgtcgatcc ccactggaaa gaccgcgaag agtttgcct 5940  
caaccgcgag ctgtggaaaa aaaaggaca ggataagtat gacatcatca agggaaaccct 6000  
ggactactgc gccctacaga tctgcagccc gggggatcca ctagttctag cctcgagtag 6060  
gaattcgata acttcgtata atgtatacta tacgaagttt tgggtcgatg gtgtatgctt 6120  
gcaattcggg tcgtatggta agcttgata acttcgtata atgtatgcta tacgaagttt 6180  
tcaattcgcc ctatagtgag tcgttattctc ccggaaaaccg cttctagcaa cgaatttcct 6240  
atactatting aagaatagga acttcggaaat aggaacttca gcagatctgc tttccattt 6300  
gcatttcaaa tgtgattttt acttacatta tacaaagctg aaaggtcagt gcaggatctg 6360  
ctttccattt tgcatttcaa attgtattttt aacttacattt atacaagct gaaaggtcag 6420  
tgcaggatct gctttccat ttgcatttca aatgtgattt taacttacattt tatacaaaacg 6480  
tggaaaggatca gtgcaggatc tgctttcca ttgcatttca aatgtgattt ttaacttac 6540  
ttatacaaaag ctgaaaggatca agtgcaggat ctgctttccat tttgcattt caaatgtgat 6600  
tttaacttac attatacaaa gctgaaaggatca cagtgccaggat tctgctttccat tttgcatt 6660  
tcaaaatgtca tttaacttac caattacaaa agtgcaggatc tcagtgccaggat atctgctt 6720  
gttcctatac ttcttagaga ataggaaactt cggaaatagga acttcgaattt ctcggggcc 6780  
cgggctcgac cagctgtcgatc catagtggct tgaatcgata aaataaaaga ttttatttt 6840  
tctccagaaaa aaggggggaa tgaaagaccc cacctgttagg tttggcaagc tagcacaacc 6900  
cctcactcgcc ggcgcggcagtc ctccgatttca ctgagtcgc cgggtaccccg tttatccat 6960  
aaaccctctt gcagttgtatca ccgacttgcgtgt ttttcgtgtt tttttttttt 7020  
gagtgttgcgtatca ctacccgtca gccccgggtct ttcacatgca gcatgtatca aaattttttt 7080  
ggttttttttt cttaagtatt tacattaaat ggccatagttt gcatatgtatca atcggccaaac 7140  
ggcgccggggaa aggccgggtttt cgtattgggc gctcttcgc ttcctcgctc actgactcg 7200  
tgcgtcgatc cgttccggatc cggcggggggatc tatcagctca ctcaaggcg gtaatacggt 7260  
tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg 7320  
ccaggaaaccg taaaaggcc gcggtcgatcc cgtttttccat taggctccgc cccccctgac 7380  
agcatcacaa aaatcgacgc tcaagtca ggtggcgaaa cccgacaggatca ctataaagat 7440  
accaggcggtt tccccctgaa agtccctcg tgcgtcgatcc tttccggacc ctgcggctt 7500

-continued

ccggataacct gtccgcctt ctcccttcgg gaagcgtggc gtttctcat agtcacgct 7560  
 gtaggtatct cagttcggtg taggtcggtc gctccaaget gggctgtgtg cacgaacccc 7620  
 ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc aaccggtaa 7680  
 gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga gcgaggtatg 7740  
 taggcggtgc tacagagttc ttgaagtggg ggcctaacta cggctacact agaaggacag 7800  
 tatttggat ctgcgtctg ctgaagccag ttaccttcgg aaaaagagtt ggtagcttt 7860  
 gatccggcaa acaaaccacc gctggtagcg gtgggttttt tgtttcaag cagcagatta 7920  
 cgccagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg tctgacgctc 7980  
 agtggAACGA aaactcacgt taagggattt tggtcatgag attatcaaaa aggatctca 8040  
 cctagatcct tttggggccg caaatcaatc taaaatatat atgagtaaac ttggtctgac 8100  
 agttacaat gcttaatcag tgaggcacct atctcagcga tctgtctatt tcgttcatcc 8160  
 atagttgcct gactccccgt cgtgttagata actacgatac gggaggggctt accatctggc 8220  
 cccagtgtcg caatgatacc gcggagaccca cgttcacccgg ctccagattt atcagcaata 8280  
 aaccagccag ccggaggccg cgagcgcaga agtggtcctg caactttatc cgcctccatc 8340  
 cagtctatta attgttgccg ggaagctaga gtaagtagt cggccagttaa tagttgcgc 8400  
 aacgttgttg ccattgtgc aggcattgtc gtgtcaecgt cgtcggtttt tatggctca 8460  
 ttcagetccg gttcccaacg atcaaggcga gttacatgat cccccatgtt gtgaaaaaaa 8520  
 ggggttagct ctttcggtcc tccgatcggtt gtcagaagta agttggccgc agtgttatca 8580  
 ctcatggtta tggcagcaact gcataattct cttactgtca tgccatccgt aagatgcttt 8640  
 tctgtgactg gtgagtagtc aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt 8700  
 tgctcttgcc cggcgtcaat acgggataat accgcgcac atagcagaac tttaaaagt 8760  
 ctcatcattt gaaaacgttc ttccgggcaaa aactctcaaa ggatcttacc gctgttgaga 8820  
 tccagttcga tgtaaccac tcgtgcaccc aactgatctt cagcatctt tactttcacc 8880  
 agcggttctg ggtgagcaaa aacaggaagg caaatgccg caaaaaaggaa aataaggcgc 8940  
 acacggaaat gttgaataact catacttttc cttttcaat attattgaag catttatcag 9000  
 ggttattgtc tcatgagccg atacatattt gaatgtattt agaaaaataa acaaataagg 9060  
 gttcccgccca cattttcccg aaaagtgcac 9091

<210> SEQ\_ID NO 6  
 <211> LENGTH: 10372  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Retroviral gene trap vector FlipROSAbetaGeoPuro

&lt;400&gt; SEQUENCE: 6

ctgcagcctg aatatgggcc aaacaggata tctgtggtaa gcagttccctg cccggctca 60  
 gggccaagaa cagatggAACGA aactgaaat gggccaaaca ggatatctgt ggttggcagt 120  
 tctgcggccg gtcaggggcc aagaacat ggtccccaga tgccggccag ccctcagcag 180  
 ttcttagaga accatcagat gttccaggg tgccccaaagg acctgaaatg accctgtgcc 240  
 ttatggatca taaccaatca gttcgcttct cgttctgtt cgcgcgttc tgctccccga 300  
 gctcaataaa agagcccaca acccctcaact cggggcgcaca gtcctccgat tgactgagtc 360  
 gcccgggtac ccgtgtatcc aataaaccct cttgcagttg catccgactt gtggtctcgc 420

-continued

---

tgttccttgg gagggctcctcc tctgagtatgat tgactaccccg tcagcgggggg tctttcattt	480
ggggggctcggt ccgggatcggt gagaccctcg cccaggggacc accgaccacac caccggggagg	540
caagctggcc agcaacttat ctgtgtctgt ccgattgtct agtgtctatg actgattta	600
tgccgcctcg tcggacttag tttagctaact agctctgtat ctggcggacc cgtgggtggaa	660
ctgacgagtt cggAACACCC ggCGGCAACC ctgggagacg tcccaggggac ttCGGGGCC	720
gtttttgtgg cccgacactga gtccaaaaat cccgategtt ttggactctt tggtgcaccc	780
cccttagagg agggatatgt gttctggta ggagacgaga acctaaaaca gttcccgcct	840
ccgtctgaat ttttgccttc gggttgggac cgaagcccgcc cccgcgcgtct tgcgtgcgc	900
agcategttc tgcgttgcct ctgtctgact gtgtttctgt atttgcgttga gaattaggc	960
cagactgtta ccactccctt aagtttgcacc tttaggtact ggaaagatgt cgagcggatc	1020
gttcacaacc agtcggtaga tgtcaagaag agacgttggg ttacccctcg ctctgcagaa	1080
tggccaacctt ttaacgtcggt atggccgcga gacggcacct ttaacccgaga cctcatcacc	1140
cagggttaaga tcaagggttcc ttccacctggc cccgcattggac acccagacca ggtccccctac	1200
atcgtgacccctt gggaaaggccctt ggcttttgcac ccccccctccctt gggtaagcc ctgttgcac	1260
cctaaggccctt cccctcccttc ttcccccattcc gccccgtctc tcccccttgc acctccctcg	1320
tcgcaccccgcc ctcgcattccctt ccccttatcca gcccctactc cttctctagg cgccggccgg	1380
atccgttgctt gaagttccata ttcccaagttt cctatttttc aaatagtata ggaacttcgt	1440
taacgaagtt cctattccgtt agtttctattt ctctagaaag tataggaact tctcggtacg	1500
gtctcgaagc cgcgggagat ctaattcttag atttaagctt ctgtatggaaat tagaacttgg	1560
caaaaacaataa ctgagaatgtt agtgtatgtt gaacagaggc tgctgtatctc gttttcagg	1620
ctatgaaactt gacacatttt gaaaccacag tacttagaaac cacaaaatgg gaatcaagag	1680
aaaaacaatgtt atcccacggat agatctatag atctatagat catgagtggg aggaatggc	1740
tggcccttaa tttttttttt cttttttttt ttatgtatccca caactatgtaa acattatcat	1800
aaagcaatag taaagagccca cagtaaagag caggcattta tcttaatccc accccacccc	1860
caccccccgtt gtcctaatcc ttccattcaa aatgttagtta ctctgtctc acccttctta	1920
acaaaatgtt acaggaaaaaa cttccatccc agtggacatc tttttttttt aatagatcat	1980
caattttctgc acacttacag cggatccctt caggcaccgg gtttgcgggtt catgcaccag	2040
gtcgccgggtt cttccggca ctcgcacgtcg gccgtgacgg tgaagccggcc cgctcgtag	2100
aagggggggtt tgccggggccggc ggagggtctcc aggaaggccgg gcaccccccggc ggcgtcgcc	2160
gcctccactc cggggagccac gacggcgctg cccagaccct tggccctgggtt gtcggccgg	2220
acgcccggccgg tggccaggaa ccacccggcc tccctggcc ggtgcggccggc caggaggcc	2280
tccatctgtt gtcggccggc cagccggaa cccgtcaactt cggccatgcg cggcccgatc	2340
tccggcgttccca cccggccggcc ttccgcgtcc tccggccgtgg tccagaccgc cccggccggc	2400
ccgtcgccggc cggccacac cttccggatc tccggccggcc cccggccgtgg gaaagatct	2460
tgcagctcggtt tgccggccgtc gatgtggccgg tccggatcgta cgggtggccgg cgtggccgggg	2520
tagtcggccgtt acggccggccggc gagggtgggtt acggccctgg ggacgtcgcc gccgggtggcc	2580
aggccgaccgg tggccgttgcata ctcggccgtc ttggctgcag gtcgaaaggc ccggagatga	2640
ggaaaggaggaa aacacggccggc cagacgtcgcc cttttgtggcc gtcgacaatg ccggccgtcc	2700
gaggaccccttcc gggccggccggc cccggccggcc agccggccccc tgagccggcc cccggaccca	2760
cccttccca gctctgtggc ccagaaacggc aaggagcaaa gtcgtatgtt gccgtggcc	2820

-continued

---

caaaggccta cccgcttcca ttgctcagcg gtgctgtcca tctgcacgag actagtgaga	2880
cgtgctactt ccatttgtca cgctctgcac gacgcgacgt gggggggggg ggggaacttc	2940
ctgactaggg gaggagtaga aggtggcgcg aagggggcac caaagaacgg agccgggttg	3000
cgccctaccgg tggatgtgga atgtgtgcga ggccagaggc cacttgtta gcgccaagtg	3060
cccagegggg ctgctaaagc gcatgctcca gactgccttggaaaagcgc ctccccctacc	3120
cggtagaatt ccccgccgtg gagcggaggaa gccggaaagagt ctagataac ttctgtatgt	3180
atacattata cgaagttatg ggtcgatgtt gagatctggaa ctagagggtc gatggtgatg	3240
cttggataac ttctgtatagc atacattata cgaagttatc ggatcccagt gtgggtgtac	3300
tcgaggtcga ctctagagga tcgagccca gctggttttt tccgtctcag aagccataga	3360
geccacccgca tccccccat gcctgttatt gtcttcccaaa tccctccccc ttgtgtctg	3420
ccccacccca ccccccagaa tagaatgaca cctactcaga caatgcgtatg caatttcctc	3480
attttatttag gaaaggacag tggggatgtgc accttccagg gtcaagggaa gcaacggggaa	3540
ggggcaaaaca acagatggct ggcaactaga aggcacagtc gaggctgatc agcgagctct	3600
agagaatttaa tccccctcaga agaactcgatc aagaaggcgaa tagaaggcgaa tgctgtgcga	3660
atcggggacgc gcgataccgt aaagcacgag gaagcggtca gcccattcgc cgccaaagctc	3720
ttcagcaata tcacgggttag ccaacgcstat gtctgtatag cggccggccca cacccagccg	3780
gccccccatcg atgaatccag aaaaggccccc atttccacc atgatattcg gcaaggaggc	3840
atcgccatgg gtcacgacga gatctcgcc gtcgggcatg cgccgccttga gcctggcgaa	3900
cagtteggct ggcggcgacc cctgtatgtc ttctgtccaga tcatcctgtat cgacaagacc	3960
ggcttccatc cgagtgacgt ctcgcgtat gcatgtttt gttttgtgtt cgaatgggca	4020
ggtagccgga tcaagcgat gcagccgcgc cattgtatca gccatgtatgg atacttttc	4080
ggcaggaggca aggtggatg acaggagatc ctgcggccgc acattcgccca atagcaggcca	4140
gtcccttccc gcttcgtatca caaagtcgag cacagctcgca caagggacgc ccgtgtggc	4200
cagccacgt agcccgctg cctgtctcg cagttcattt cggccaccgg acagggtcggt	4260
cttgacaaaa agaaccgggc gccccctgcgc tgacagccgg aacacggccg catcagagca	4320
gcccgttgc tttttgtgtccc agtcatatgc gaatacgctc tccacccaag cggccggaga	4380
acctgtgtgc aatccatctt gttcaatggc cgatcccata ttggctgcag cccggggat	4440
ccccctgacac cagaccacact ggtatggta gggccggcc ctcagcttgcgaa attccgcga	4500
tactgacggg ctccaggagt cgtcgccacc aatccccata tggaaaccgt cgatattcg	4560
ccatgtgcct tttcccgctg gcagcagatg gcatggctg gttccatca ttgtgtttt	4620
actgtgtcggt ctgtatgttgc actggaaatgc gcccggccac tgggtgtggc cataattcaa	4680
ttcgccgcgc ccccgccgc gaccgtttt gctcggttgc acgtacgggg tatacatgtc	4740
tgacaatggc agatccccatc ggtcaaaaaca ggccggcgttca aggccggccgg gatagttt	4800
ttcgccgcct aatcccgagcc agtttaccgg ctctgttacc tgcggccagct ggcagttcag	4860
gccaatccgc gcccggatgcg gtgtatcgat cgccacttca acatcaacgg taatcgccat	4920
ttgaccacta ccatcaatcc ggttaggtttt ccggctgtata aataagggtt tccccctgtat	4980
ctgcccacgc tgagcggtcg taatcagcac cgcatcagca agtgttatctg ccgtgcactg	5040
caacaacgcgt gcttcggccct ggtatggcc cggccgccttc cagcgttgcgaa cccaggcggtt	5100
agggtcaatg cgggtcgctt cacttacgccc aatgtcgat tccagcggtg cacgggtgaa	5160

-continued

---

ctgatecgcbc agcggcgtca gcagttgttt tttatcgcca atccacatct gtgaaagaaa	5220
gcctgactgg cggtaaattt gccaacgctt attaccaggc tcatgtcaaa aatccatcc	5280
gctgggtggc agatgcggga tggegtggga cgccgggggg agcgtcacac tgaggtttc	5340
cgcacagcgc cactgtgcc aggccgtat gtgcgggct tctgaccatg cggtegcgtt	5400
cggttgact acgcgtactg tgccgcagag ttgcccggc ctctccggc gccgttagttc	5460
aggcagtca atcaactgtt tacattgtgg agcgacatcc agaggcactt caccgctgc	5520
cagcggctta ccatccagcg ccaccatcca gtgcaggagc tcgttatcgc tatgacggaa	5580
caggtattcg ctggtaactt cgatggttt cccggataaa cggaaactgga aaaactgctg	5640
ctgggtttt gctccgtca ggcgtggatg cggcgtgcgg tcggcaaaga ccagaccgtt	5700
catacagaac tggcgtatgt tcggcgtatc gccaaaatca ccggcgtaag ccgaccacgg	5760
gttgcgtttt tcatcatatt taatcagcga ctgatccacc cagtcggcaga cgaagccgcc	5820
ctgtaaacgg ggatactgac gaaacgcctg ccagtttta gcgaaaccgc caagactgtt	5880
acccatcgcg tggcgtatt cgaaaggat cagcggcgcg gtctctccag gtagcggaaag	5940
ccatTTTTCG atggaccatt tcggcacagc cgggaagggc tggctttcat ccacgcgcgc	6000
gtacatcggtt caataatata cggtggccgt ggtgtcggtt ccggccgtt catactgcac	6060
cgggcgggaa ggatcgacag atttgatcca gcgatacagc gcgtcggtat tagcgcgtg	6120
gcctgattca ttccccagcg accagatgtt cacactcggtt tgattacgtat cgcgtgcac	6180
cattcgcgtt acgcgttcgc tcatcgccgg tagccagcgc ggatcatcggtt tcagacgtt	6240
cattggcacc atgcccgtggg tttcaatatt ggcttcatcc accacataca ggccgttagcg	6300
gtcgcacagc gtgtaccaca gcggatgggtt cggataatgc gaacagcgcgaa cggcgtaaa	6360
gttgcgttcgc tcatcagca ggatccctgc caccatcgatc tgctcatccat tgacctgacc	6420
atgcagagga ttagtgcgtt gacggtaac gcctcgaatc agcaacggct tgccgttcag	6480
cagcagcaga ccatttcaaa tccgcacccgc gggaaaccgc acatcgccgg ctctcgcttc	6540
aatcagcgtt cgcgtggcggt tgcgttgtt aaccaccgcgat cgatagagat tcgggatttc	6600
ggcgctccac agtttccgggtt ttgcgttgtt cagacgttgtt gtgcgtcgat cggcataacc	6660
accacgctca tcgataattt caccggcggaa aggcgcgggtt cgcgtggcgat cctgcgttcc	6720
accctgcctt aaagaaactg ttaccgttag ttagtcacgc aactcgccgc acatctgaac	6780
ttagcgttcgc agtacagcgc ggcgtaaatc atcattaaatcg agtggccaa catggaaatc	6840
gtgttgcgtt ttagtgcgtt cgcgttgtt cgcgttgtt cgcgttgtt cgcgttgtt cgcgttgtt	6900
ccacatatcc ttagtgcgtt cgcgttgtt cgcgttgtt cgcgttgtt cgcgttgtt cgcgttgtt	6960
gggggtttttt cccggcgatc aaaatgcgtt caggtcaat tcaagacggca aacgactgtc	7020
ctggccgtaa cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7080
aaaaataattt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7140
gttgcgttgtt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7200
ttagtgcgtt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7260
atgtatcggtt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7320
aaaccaggca aagcgccatt cgcgttgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7380
gggggtttttt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7440
ttgggttaacg ccagggtttt cccggcgatc gggccgttgcgatcc cccggcgatc aacgcccgttgcgatcc	7500
acagatcatc aagcttacgc ataccgtcga tccccactgg aaagaccgcgaa aagagttgtt	7560

-continued

---

cctcaaccgc gagctgtgga aaaaaaaggg acaggataag tatgacatca tcaaggaaac 7620  
 cctggactac tgcgcctac agatctgcag cccggggat ccactagtgc tagcctcgag 7680  
 taggaattcg ataacttcgt ataatgtata ctatacgaag ttatgggtcg atggtgatgc 7740  
 ttggcaattc gggtcgatgg tgaagcttgg ataacttcgt ataatgtatg ctatacgaag 7800  
 ttatcaattc gccctatagt gagtcgtatt ctcccggaaa ccgcttctag caacgaagtt 7860  
 cctatactat ttgaagaata ggaacttcgg aataggaact tcagctctgt atcgtctatc 7920  
 gttatgtcta atggtatctg atgcagatct tctgtatctgt ctatcggtat gtctaattgt 7980  
 atctgtatc gatctgtatc agttccata ctttcttagag aataggaact tcggaatagg 8040  
 aacttcgaat tctcgagggc ccgggctcga ccagctgtgc gcatagtggc ttgaatcgat 8100  
 aaaaataaaag attttattta gtctccagaa aaagggggggg atgaaagacc ccacccgttag 8160  
 gtttggcaag ctatcacaaac ccctcactcg gggcgccagt cctccgattt actgagtcgc 8220  
 ccgggtatccc gtgtatccaa taaaccctct tgcagttgc tccgacttgt ggtctcgctg 8280  
 ttcccttggga gggctctctc tgagtgattt actaccggc agcgggggtc tttcacatgc 8340  
 agcatgtatc aaaattaatt tggtttttt tcttaagtat ttacattaa tggccatagt 8400  
 tgcattaaatg aatcgccaa cgccggggga gaggccgtt ggttattggg cgctttccg 8460  
 cttccctcgct cactgactcg ctgcgcctgg tgcgttgcgtt gggcgagcg gtatcgctc 8520  
 actcaaaggc ggtaataacgg ttatccacag aatcaggggg taacgcagga aagaacatgt 8580  
 gagccaaaagg ccageaaaag gccaggaacc gtaaaaaggc cgccgttgcg gcgttttcc 8640  
 ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtctg aggtggcgaa 8700  
 acccgacagg actataaaga taccaggcg tttccctgg aagctccctc gtgcgcctc 8760  
 ctgttccgac cctgcgcgtt accggatacc tgcgcgtt tctcccttc ggaagcggtt 8820  
 cgctttctca tagctcacgc tgcgttgcgtt tgcgttgcgtt cgctccaagc 8880  
 tgggtgtgtt gcacgaaccc cccgttcgcg ccgaccgtcg cgccattatcc ggtaactatc 8940  
 gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 9000  
 ggattagcag agcgaggtat gtaggcggtt ctacagagtt ctgaaagtgg tggcctaact 9060  
 acggctacac tagaaggaca gtattggta tctgcgttgcgtt gtcgaagccca gttacccctc 9120  
 gaaaaaaggt tggtagctt tgcgttgcgtt aacaaaccac cgctggtagc ggtgggtttt 9180  
 ttgtttgcgtt gcacgttgcgtt acggcgatggaaa aaaaaggatc tcaagaagat cttttgtatct 9240  
 ttcttacggg gtctgacgtt cagttggaaacg aaaactcactt ttaagggtt ttgggtcatga 9300  
 gattatcaa aaggatcttcc acctagatcc ttttgcggcc gcaaatcaat cttaagtata 9360  
 tatgagtaaa ctgttgcgtt cagttacca tgcgttgcgtt gtcgttgcgtt tatctcagcg 9420  
 atctgtctat ttgcgttgcgtt cagttggaaacg aaaactcactt ttaagggtt ttgggtcatga 9480  
 cgggggggtt taccatctgg ccccgatgtgtt gcaatgatacc cgccggggcc acgctccaccg 9540  
 gtcggatgtt tatcgcaat aaaccagccca gcccggaaaggcccgccggcccg aagtgggtt 9600  
 gcaactttat cccgttccat ccgttgcgtt aattttttttt gggaaagctttagt agtaatgtt 9660  
 tcggccgttta atagtttgcgtt caacgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 9720  
 tcgttgcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 9780  
 tcccccattgt tgcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt gtcgttgcgtt 9840  
 aagttggccg cagttttatc actcatggttt atggcagcac tgcataattc tcttactgtc 9900

-continued

---

atgccccatccg	taagatgctt	ttctgtgact	ggtgagtaact	caaccaaggc	attctgagaa	9960
tagtgttatgc	ggcgaccgag	ttgtcttgc	ccggcgtcaa	tacgggataa	taccgcgcca	10020
catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacgtt	cttcggggcg	aaaactctca	10080
aggatcttac	cgctgttgag	atccagttcg	atgtaaacca	ctcggtcacc	caactgatct	10140
tcagcatctt	ttactttcac	cagcggttct	gggtgagca	aaacaggaa	gcaaaatgcc	10200
gaaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcatactctt	ccttttcaa	10260
tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	gatacatatt	tgaatgtatt	10320
tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	gaaaagtgcc	ac	10372

&lt;210&gt; SEQ ID NO 7

&lt;211&gt; LENGTH: 10501

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial

&lt;220&gt; FEATURE:

<223> OTHER INFORMATION: Retroviral gene trap vector  
eFlip3ROSAbetaGeoPuro

&lt;400&gt; SEQUENCE: 7

ctgcagccgt	aatatgggcc	aaacaggata	tctgtggtaa	gcagttccgt	ccccggctca	60
gggccaagaa	cagatggAAC	agctgaatat	ggggccaaaca	ggatatctgt	ggtaaggcagt	120
tcctgcggcc	gctcaggccc	aagaacagat	ggtccccaga	tgccggccag	ccctcagcag	180
tttcttagaga	accatcagat	gtttccaggg	tgcggcaagg	acctgaaatg	accctgtgcc	240
tttatttgaac	taaccaatca	gttgcgttct	cgcttctgtt	cgccgcgttc	tgctccccga	300
gctcaataaa	agageccaca	accctcact	cgggggcgc	gtccctcgat	tgactgagtc	360
gccccgggtac	ccgtgtatcc	aataaaccc	cttgcagttt	catccgactt	gtggtctcgc	420
tgttccttgg	gagggtctcc	tctgagtgat	tgactacccg	tcagcggggg	tcttcattt	480
ggggggctcg	ccgggatcg	gagaccctg	cccaggacc	accgaccac	caccggagg	540
caagctggcc	agcaacttat	ctgtgtctgt	ccgattgtct	agtgtctatg	actgatttt	600
tgcgcctcg	tgggtactag	ttagctaact	agctctgtat	ctggccggacc	cgtggtgaa	660
ctgacgagtt	cggaacaccc	ggccgcaccc	ctgggagacg	tcccaaggac	ttcgggggcc	720
gtttttgtgg	cccgacctga	gtccaaaat	cccgatgtt	ttggactctt	tggtgcaccc	780
cccttagagg	agggatatgt	ggttctggta	ggagacgaga	acctaaaaca	gttcccgcct	840
ccgtctgaat	ttttgttttc	ggtttggac	cgaaagccgc	ccgcgcgtct	tgtctgtgc	900
agcategttc	tgtgttgtct	ctgtctgact	gtgtttctgt	atttgtctga	gaattaggc	960
cagactgtta	ccactccctt	aagttgacc	ttaggtcaact	ggaaagatgt	cgagcggatc	1020
gtctcacaacc	agtcggtaga	tgtcaagaag	agacgttggg	ttacctctg	ctctgcagaa	1080
tggccaaacct	ttaacgtcg	atggccgcga	gacggcacct	ttaaccgaga	cctcatcacc	1140
caggtaaga	tcaaggctt	ttcacctggc	ccgcacggac	acccagacca	ggtcccctac	1200
atcgtgacct	gggaaggcctt	ggcttttgc	ccccctccct	gggtcaagcc	ctttgtacac	1260
cctaagccctc	cgccctccct	tcctccatcc	ccccgggtcc	tcccccgttga	accttctcg	1320
tcgacccgc	ctcgatccctc	cctttatcca	gcctcactc	cttctctagg	cgccggccgg	1380
atccgttgct	gaagttccta	ttccgaagtt	cctattttcc	aaatagata	ggaacttcgt	1440
taacgaagtt	cctattccga	agttcctatt	ctctagaaag	tataggaact	tctcggtacg	1500
gtctcgaagc	cgcggagat	ctaattctag	attnaagttt	ctgtatggaa	tagaacttgg	1560

-continued

---

caaaaacaata	ctgagaatga	agtgttatgt	gaacagaggc	tgctgatctc	gttcttcagg	1620
ctatgaaact	gacacatttg	gaaaccacag	tacttagaac	cacaaaagtgg	gaatcaagag	1680
aaaaacaatg	atcccacgag	agatctatag	atctatagat	catgagtggg	aggaatgagc	1740
tggcccttaa	tttggtttag	cttggtaaaa	ttatgatatac	caactatgaa	acattatcat	1800
aaggcaatag	taaagagcca	cagtaaagag	cagggcattt	tcttaatccc	accccacccc	1860
cacccccgt	gctccaatcc	ttccattcaa	aatgttaggt	ctctgttctc	acccttctta	1920
acaaaagtatg	acaggaaaaaa	cttccatTTT	agtggacatc	tttattgttt	aatagatcat	1980
caatttctgc	agacttacag	cggatcccct	cagggcaccgg	gcttgegggt	catgcacca	2040
gtcgcgcggt	ccttcgggca	ctcgacgtcg	gccccgtacgg	tgaagccgag	ccgctcgtag	2100
aaggggaggt	tgccccggc	ggagggtctcc	aggaaggccg	gcacccccc	gccccggcc	2160
gcctccactc	cggggagcac	gacggcgctg	cccagacccct	tgccctgggt	gtcggggcag	2220
acgcccacgg	tggccaggaa	ccacccgggc	tccttgggccc	ggtgcggccg	caggaggcct	2280
tccatctgtt	gtcgccggc	cagccgggaa	ccgctcaact	cgccatgcg	cgccatgcgatc	2340
tcggcgaaca	ccgccccccgc	ttcgacgctc	tccggcgtgg	tccagaccgc	caccggccgc	2400
ccgtcgcccg	cgacccacac	cttgcgcgt	tcgagccccg	cgccgtcgag	gaagagttct	2460
tgcaagctcg	tgaccccgctc	gatgtggccg	tccggatcga	cggtgtggcg	cgtggccggg	2520
tagtcggcga	acgcggccgc	gagggtgcgt	acggccctgg	ggacgtcg	ccgggtggcg	2580
aggcgcacccg	tgccgttgc	ctcggtcata	ttggctgcag	gtcgaaaggc	ccggagatga	2640
ggaagaggag	aacagcgccg	cagacgtcg	cttttgaagc	gtgcagaatg	ccgggtctccg	2700
gaggaccttc	gggcgcccgc	cccgccccctg	agccgcggcc	ttagcccgcc	cccgaccca	2760
cccccttccca	gcctctgagc	ccagaaagcg	aaggagcaa	gtgtctatg	gccgtgccc	2820
caaaggccta	cccgcttcca	ttgctcagcg	gtgtgttcca	tctgcacgag	actagtgaga	2880
cgtgctactt	ccatttgtca	cgtctgcac	gacgcgacgt	gcggggccgg	ggggaaacttc	2940
ctgactaggg	gaggagtaga	aggtggccgc	aaggggccac	caaagaacgg	agccgggttg	3000
cgcctaccgg	tggatgttga	atgtgtgcga	ggccagaggc	cacttgttga	gcccggatgt	3060
cccaagccggg	ctgctaaagc	gcatgctcca	gactgccttgc	ggaaaagcgc	ctcccttacc	3120
cggtagaatt	cccccgccgt	gagcgaggaa	gcggaaagagt	ctagaataac	ttcgtatagt	3180
atacattata	cgaagttatg	ggtcgtatgt	gagatctgg	ctagagggtc	gatggtgatg	3240
cttggataac	ttcgatatac	atacattata	cgaagttatc	ggatccca	gtgggtggtac	3300
tcgagggtcga	ctcttagagga	tcgagccca	gctggtttt	tccgtctcag	aagccataga	3360
geccaccgc	tccccagcat	gcctgttatt	gtttttccaa	tccctccctt	tgctgtctcg	3420
ccccacccca	ccccccagaa	tagaatgaca	cctactcaga	caatgcgtat	caatttctc	3480
attttattag	gaaaggacag	tgggatgtgc	actttccagg	gtcaaggaag	gcacggggga	3540
ggggcaaaaca	acagatggct	ggcaactaga	aggcacatgc	gaggctgtatc	agcgagctct	3600
agagaattga	tccccctcaga	agaactcg	aagaaggcga	tagaaggcga	tgcgtcg	3660
atcggggagcg	gcgataaccgt	aaagcacag	gaagcggtca	gcccattcgc	cgccaaagctc	3720
ttcagaata	tcacgggtat	ccaaacgtat	gtcctgtat	cggtccgc	cacccagccg	3780
gcacacgtcg	atgaatccag	aaaagcgcc	atttccacc	atgtatattcg	gcaaggcaggc	3840
atcgccatgg	gtcaacgacga	gatectcgcc	gtcgggc	cgccgttga	gcctggcgaa	3900
cagttcggt	ggcgccgagcc	cctgtatgc	ttcgtccaga	tcatctgtat	cgacaagacc	3960

-continued

-continued

---

gtcgacacgc gtgtaccaca gcgatggtt cgataatgc gaacagcgcg cggcgtaaa	6360
gttgttgc ttcatcagca ggatatctg caccatcg tgcgtatcca tgacctgacc	6420
atgcagagga ttagtgcgt gacggtaac gcctcgaatc agcaacggct tgccgttc	6480
cagcagcaga ccatttcaa tccgcaccc gggaaaccg acatcgagg ctctgc	6540
aatcagcgtg cccgtgggg tggcgttc aaccaccgca cgatagagat tcgggat	6600
ggcgctccac agtttccgggt ttgcgtt cagacgtgt gtgcgcgt cggcataacc	6660
accacgctca tcgataattt caccgcgaa aggccgggtg cccgtggcga cctgcgtt	6720
accctgcatt aaagaaaactg ttaccgtt gtagtcacgc aactcgccgc acatctgaac	6780
ttcagcctcc agtacagcgc ggctgaaatc atcattaaag cgagtgccaa catggaaatc	6840
gtgttgcgtt gtagtgcgtt tatgcagca cggacgtca cggaaaatgc cgctcatcc	6900
ccacatatcc tgatcttcca gataactgcc gtcactccaa cgcagcacca tcaccgcgag	6960
ggggtttctt ccggcgcgtt aaaaatgcgt caggtcaat tcagacggca aacgactgtc	7020
ctggccgtaa ccgaccgcg cccgttgcg ccacagatga aacgcccagt taacgcattc	7080
aaaaataatt cgcgtctggc cttctgtt ccagcttca tcaacattaa atgtgagcga	7140
gtacaacccc gtccgttcc cccgtggaaac aaacggcggta ttgaccgtaa tgggataggt	7200
tacgttgcgtt tagatggcg catcgtaacc gtgcattgc cagtttggagg ggacgcac	7260
agtatcgcc tcaggaagat cgcactccag ccagcttcc ggcaccgcgtt ctggcgcgg	7320
aaaccaggca aagccgcatt cgcattcag gctgcgcac tggtggaaag ggcatcggt	7380
gcgggcctct tcgcttattac gccagctggc gaaagggggaa tggctgcac ggcgatgg	7440
ttggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggat cccatgtc	7500
acagatcatc aagcttatcg ataccgtcg tccccactgg aaagaccgcg aagagttgt	7560
cctcaaccgc gagctgtgg aaaaaaggg acaggataag tatgacatca tcaaggaaac	7620
cctggactac tgcgcctac agatctgcacg cccggggat ccactagtcc tagctcgag	7680
taggaattcg ataacttcgt ataatgtata ctatacgaag ttatgggtcg atgggtatgc	7740
ttggcaattc gggtcgtatgg tgaagcttgg ataacttcgt ataatgtatg ctatacgaag	7800
ttatcaattc gcccatagt ggtcgattt cttccggaaa cccgttctag caacgaagtt	7860
cctataactat ttgaagaata ggaacttcgg aataggaact tcagcagatc tgctttcca	7920
tttgcatttc aaatgtgatt ttaacttaca ttatacaaag ctgaaaggc agtgcaggat	7980
ctgctttcc atttgcattt caaatgtat tttaacttac attatacaa gctgaaagg	8040
cagtcaggg tctgttcattt catttcattt tcaaatgtat tttaacttac cattatacaa	8100
agtcgaaagg tcagtcagg atctgcgtt gttcctatac ttcttagaga ataggaaat	8160
cgaaatagga acttcgaatt ctcgaggccc cgggctcgac cagctgtgcg catagtggct	8220
tgaatcgata aaataaaaga ttttatgg tctccggaaa aaggggggaa tgaaagaccc	8280
cacctgttagg ttggcaagc tagcacaacc cctcaactcg ggcgcgcgtc ctccgat	8340
ctgagtcgcg cgggtacccg tggatccat aaaccctctt gcagttgcgt ccgacttgc	8400
gtctcgctgt tccttggag ggtctccctt ggttgcgttca cggggggct	8460
ttcacatgca gcatgtatca aaattaattt ggttttttt cttaagtatt tacattaaat	8520
ggccatgtt gcatatgttga atcggccaaac ggcggggggagg aggccgttgcgtt cgtattggc	8580
gtcttcgcg tccctcgctc actgactcg tgcgttcggt cttccgtc cggcgacgg	8640
tatcagtcgtca ctcaaaaggcg gtaatacggt tatccacaga atcagggtt aacgcaggaa	8700

-continued

agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgcgg	8760
cgttttcca taggctccgc cccccctgacg agcatcacaa aaatcgacgc tcaagtcaaga	8820
ggtggcgaaa cccgacagga ctataaagat accaggcggtt tccccctgga agctccctcg	8880
tgcgctctcc tggccgacc ctggccgctta ccggataacct gtccgectt ctcccttcgg	8940
gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcggtc	9000
gctccaagct gggctgtgtg cacgaacccc cggttcagcc cgaccgctgc gccttatccg	9060
gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca	9120
ctggtaacag gattagcaga gcgaggtatg taggcgggtgc tacagagtcc ttgaagtgg	9180
ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag	9240
ttaccttcgg aaaaagagtt ggtagcttt gatccggcaaa acaaaccacc gctggtagcg	9300
gtggttttt tggggcaag cagcagatta cgccgcgaaa aaaaggatct caagaagatc	9360
ctttgatctt ttctacgggg tctgacgctc agtggAACAG aaactcacgt taagggattt	9420
tggcatgag attatcaaaa aggatcttca cctagatctt tttgcggccg caaatcaatc	9480
taaagtatat atgagtaaac ttgggtctgac agttaccaat gttaatcag tgaggcacct	9540
atctcagcga tctgttattt tcgttcatcc atagttgcctt gactccccgt cgttagata	9600
actacgatac gggagggctt accatctggc cccagtgctg caatgatacc gcgagaccca	9660
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaaaggcc cgagegcaga	9720
agtggctctg caactttatc cgcctccatc cagtcttattt attgttgcgg ggaagctaga	9780
gtaagtagtt cgccagttaa tagttgcgc aacgttgcgtt ccattgtgc aggcatcg	9840
gtgtcacgt ctgcgtttgg tatggcttca ttcaagctccg gttcccaacg atcaaggcga	9900
gttacatgtat ccccatgtt gtcaaaaaaa gcggttagct ctttcgggtcc tccgatcg	9960
gtcagaagta agttggccgc agtgttatca ctcattgttca tggcagactc gcataattct	10020
cttactgtca tgccatccgt aagatgtttt tctgtgactg gtgagttactc aaccaagtc	10080
ttctgagaat agtgttatgcg gcgaccgagt tgctcttgcg cggcgtcaat acgggataat	10140
accgcgccac atagcagaac tttaaaagtg ctcatcattt gaaaacgttc ttccggcga	10200
aaactctcaa ggatcttacc gctgttgaga tccagttcgta tgtaaccac tcgtgcaccc	10260
aactgatctt cagcatctt tactttcacc agcggttctg ggtgagcaaa aacaggaagg	10320
caaaatgcgg caaaaaaggaa aataaggccg acacggaaat gttgaataact catactttc	10380
ctttttcaat attattgaag catttacag ggttattgtc tcatgagccg atacatattt	10440
gaatgttattt agaaaaataa acaaataggg gttccgcgc aatttccccg aaaagtgc	10500
c	10501

```

<210> SEQ_ID NO 8
<211> LENGTH: 10710
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: Retroviral gene trap vector
    eFlip6ROSAbetaGeoPuro

<400> SEQUENCE: 8

ctgcagcgtt aatatggggc aaacaggata tctgtggtaa gcaagttctg ccccggtca      60
gggccaagaa cagatggaaac agctgaatat gggccaaaca ggatatctgt ggttaaggcagt   120
tcctgccccg gctcaggggcc aagaacagat ggtccccaga tgcggtccag ccctcagcag   180

```

-continued

-continued

---

gtcggcgaac	gccccggcgaa	gggtgcgtac	ggccctgggg	acgtcgctgc	gggtggcgag	2580
gcccacccgt	ggcttact	cggcatatt	ggctgcagg	cgaaaggccc	ggagatgagg	2640
aaggaggaga	cagcgccgca	gacgtgcgt	tttgaagcgt	gcagaatgcc	gggctccgga	2700
ggacacctcg	gccccggccc	cggccctgag	ccccccctg	agcccccccc	cggaccacc	2760
ccttcccagc	ctctgagccc	agaaagcgaa	ggagcaaagc	tgctattggc	cgctgcccc	2820
aaggccattacc	cgcttccatt	gctcagcggt	gtgtccatc	tgcacgagac	tagtgagacg	2880
tgctacttcc	atttgtcacg	tcctgcacga	cgcgagctgc	ggggcggggg	ggaacttcc	2940
gactaggggg	ggagtagaag	gtggcgcgaa	ggggccacca	aagaacggag	ccgggtggcg	3000
cctaccgggt	gatgtggaat	gtgtgcgagg	ccagaggcca	cttgtgtagc	gccaagtgcc	3060
cagcggggct	gctaaagcgc	atgctccaga	ctgccttggg	aaaagcgct	ccccctaccc	3120
gtagaattcc	ccgggtggagc	gaggaagcgg	aagagtctag	aataacttcg	tatagtatac	3180
attatacga	gttatgggtc	gatggtgaga	tctggactag	agggtcgatg	gtgtgcctg	3240
gataacttcg	tatagcatac	attatacga	gttatcgat	cccagtgtgg	ttgtactcga	3300
ggtcgactct	agaggatcga	gccccagctg	gttctttccg	tctcagaagc	catagagccc	3360
accgcattccc	cagcatgcct	gctattgtct	tcccaatcc	cccccttgc	gtcctgcccc	3420
accccaaaaa	ccagaataga	atgacaccta	ctcagacaat	gcgtatgc	atccctcatt	3480
tattaggaaa	ggacagtggg	agtggcacct	tccagggtca	aggaaggcac	gggggagggg	3540
caaacaacag	atggctggca	actagaaggc	acagtgcagg	ctgatcagcg	agctctagag	3600
aattgatccc	ctcagaagaa	ctcgtcaaga	aggcgataga	aggcgatgc	ctgcgaatcg	3660
ggagcggcga	taccgtaaag	cacgaggaag	cggtcagccc	attcgcgc	aagctttca	3720
gcaatatac	gggtagccaa	cgctatgtcc	tgatagcggt	ccgcccaccc	cagccggcca	3780
cagtcgtat	atccagaaaa	cgcccaattt	tccaccatg	tattcggcaa	gcaggcatcg	3840
ccatgggtca	cgacgagatc	ctcgccgtcg	ggcatgcgc	ccttgcgc	ggcgaacagt	3900
tccgtggcg	cgagccccctg	atgctttcg	tccagatcat	cctgtatgc	aagaccgc	3960
tccatccgag	tacgtgtctcg	ctcgatgcga	tgtttcg	ggtggcgaa	tggcaggt	4020
gcggatcaa	gcgtatgcag	ccggccatt	gcatcagcca	tgtggatac	tttctggca	4080
ggagcaagg	gagatgacag	gagatcctgc	ccggcaact	cgcccaatag	cagccagtcc	4140
cttccgcctt	cagtgacaac	gtcgagcaca	gctgcgc	gaacgcccgt	cgtggccagc	4200
cacgatagcc	gcgtcgctc	gtcctgcagt	tcattcagg	cacggacag	gtcggcttg	4260
acaaaaagaa	ccggcgcccc	ctcgctgac	agccggaaca	cgccggcatc	agagcagccg	4320
attgtctgtt	gtccccagtc	atagccaat	agectctcc	cccaagcg	cgggaaacct	4380
gcgtcaatc	catcttgc	aatggccat	cccatattgg	ctgcagcc	ggggatcccc	4440
tgacaccaga	ccaactggta	atggtagcga	ccggcgctca	gtggaaattc	cgccgatact	4500
gacgggctcc	aggagtgcgtc	gccaccaatc	cccatatgg	aaccgtcgat	attcagccat	4560
gtgccttc	ccgcgtgcag	cagatggcga	tggctgg	ccatcagt	ctgttgc	4620
tagcggtc	tgttgaactg	gaagtcgc	cgcactgg	gtggccata	attcaattcg	4680
cgcgtccgc	agcgcagacc	gtttcgctc	ggaaagacgt	acggggata	catgtctgac	4740
aatggcagat	cccgccgtc	aaaacaggcg	gcagtaaggc	ggtcgccata	gtttctgc	4800
ggccctaatac	cgagccagtt	taccgcgtct	gctacctgc	ccagctggca	gttcaggcca	4860
atcccgcccg	gatgcgggt	atcgctgc	acttcaacat	caacggtaat	cgccatttga	4920

-continued

---

ccactaccat caatccggta ggtttccgg ctgataaaata aggtttccc ctgatgctgc 4980  
 cacgcgtgag cgggtcgtaat cagcacccga tcagcaagt tatctgcgt gcactgcaac 5040  
 aacgctgctt cggcctggta atggcccgc gccttcage gttcgaccca ggcgttaggg 5100  
 tcaatgcggg tcgcttcaact tacgccaatg tcgttatcca ggggtgcacg ggtgaactga 5160  
 tcgcgcagcg gcgtcagcag ttgttttta tcgccaatcc acatctgtga aagaaagcct 5220  
 gactggcggt taaattgcca acgcttatta cccagctcga tgcaaaaatc catttcgctg 5280  
 gtggtcagat gcgggatggc gtgggacgcg gccccggacgc tcacactgag gttttccgc 5340  
 agacgcacact gctgecagggc gctgtatgtc cgggttcgt accatgcggg cgcggttcgg 5400  
 tgcactacgc gtactgtgag ccagagttgc cggcgctct cggcgctgccc tagttcaggc 5460  
 agttcaatca actgttttacc ttgtggagcg acatccagag gcacttcacc gtttgcacgc 5520  
 ggcttaccat ccagegccac catccagtgc aggagctcgt tatcgctatg acggaacagg 5580  
 tattcgttgg tcacttcgat ggtttcccg gataaaacggg actggaaaaa ctgctgttgg 5640  
 tgttttgtt ccgtcagcgc tggatgcggc gtgcggtcgg caaagaccag accgttcata 5700  
 cagaactggc gatcggttcgg cgtatcgca aaatcaccgc cgtaagccga ccacgggttg 5760  
 ccgtttcat catatTTTaat cagcgactga tccacccagt ccacagacgaa gccgcctgt 5820  
 aaacggggat actgacgaaa cgcctgccag tatttagcga aaccgcgaag actgttaccc 5880  
 atcgcgttggg cgtattcgca aaggatcagc gggcgctct ctccaggttag cgaaagccat 5940  
 tttttgtgg accatttcgg cacagccggg aagggtcggtt cttcatccac ggcgcgtac 6000  
 atcgggcaaa taatatcggt ggccgtgggt tggctccgc cgccttcata ctgcaccggg 6060  
 cgggaaggat cgacagattt gatccagcga tacagcgcgt cgtgatttagc ggcgtggcct 6120  
 gattcattcc ccagegacca gatgatcaca ctccgggtat tacgatcgcc ctgcaccatt 6180  
 cgcgttacgc gttcgtcat cgcggtagc cagcgcggat catcggtcag acgattcatt 6240  
 ggacccatgc cgtgggttc aatattgggt tcatccacca catacaggcc gtagggcgt 6300  
 cacagegtgt accacagcgg atgggtcgga taatgcgaac agcgcacggc gttaaagttg 6360  
 ttctgtttca tcagcaggat atcctgcacc atcgtctgtc catccatgac ctgaccatgc 6420  
 agaggatgt gtcgtgacg gttaacgcct cgaatcagca acggcttgc gttcagcagc 6480  
 agcagaccat ttcaatccg cacctcgggg aaaccgcacat cgcaggctc tgcttcaatc 6540  
 agcgtgcgtt cggcggtgtg cagttcaacc accgcacgat agagattcgg gatttcggcg 6600  
 ctccacagtt tcgggttttc gacggtcaga cgtagttgtga cgcgatcgcc ataaccacca 6660  
 cgctcatcgta taatTCacc gccgaaaggc ggggtgcggc tggcgaccccg cgtttcaccc 6720  
 tggccataaag aaactgttac ccgttaggtat tcacgcacact cgccgcacat ctgaacttca 6780  
 gcctccagta cagcgccgct gaaatcatca ttaaagcggg tggcaacatg gaaatcgctg 6840  
 atttgtgttag tcgggtttatg cagcaacgag acgtcacggaa atatgcgcct catccgcac 6900  
 atatcctgtat ttccagata actgcccgtca ctccaaacgcga gcaccatcac cgcgaggcgg 6960  
 ttttctccgg cgcgtaaaaa tgcgtcagg tcaaattcag acggcaacacg actgtccctgg 7020  
 ccgtaaaccga cccagcggcc gttgcaccc acgtgaaacg ccgagttaac gccatcaaaa 7080  
 ataattcgcg tctggccttc ctgttagccag ctttcatcaa cattaaatgt gagcgagtaa 7140  
 caacccgtcg gatttccgtt gggaaacaaac ggcggattga ccgtaatggg ataggttacg 7200  
 ttgggtttaga tggggccatc gtaaccgtgc atctgcccagt ttgaggggac gacgacagta 7260

-continued

---

tccggcctcag gaagatcgca ctccagccag ctttccggca ccgcttctgg tgccggaaac	7320
caggcaaagc gccattcgcc attcaggctg cgcaactgtt gggaaaggcg atcggtgccg	7380
gcctcttcgc tattacgcca gctggcgaaa gggggatgtg ctgcaaggcg attaagtgg	7440
gtaacgcacgg gttttccca gtcacgacgt tgtaaaacga cgggatccgc catgtcacag	7500
atcatcaagc ttatcgatac cgtcgatccc cactggaaag accgcgaaga gtttgtctc	7560
aaccgcggacg tgtggaaaaa aaaggggacag gataagtatg acatcatcaa ggaaacctg	7620
gactactgcg ccctacagat ctgcagcccg ggggatccac tagttctagc ctgcagtagg	7680
aattcgtataa cttcgtataa tgtatactat acgaagtat gggtcgatgg ttagtgcgttgg	7740
caattcgggt cgatgggtgaa gcttggataa cttcgtataa tgtatgttat acgaagtat	7800
caattcgcccc tatagtgtgt cgtattctcc cgaaaaccgc ttctagcaac gaaggctcta	7860
tactatttga agaataggaa cttcggataa ggaacttcag cagatctgtt tttccatttg	7920
catttc当地 gtgattttaa cttacattat acaaagctga aaggctcgtt caggatctgc	7980
ttttccattt gcatttcaaa tgtgattttta acttacatta tacaaaagctg aaggctcgt	8040
gcaggatctg cttttccatt tgcatttcaaa atgtgattttt aacttacattt atacaagct	8100
gaaaggctcgttgcaggatct gctttccat ttgcatttca aatgtgattt taacttacat	8160
tatacaaagc tgaaagggtca gtgcaggatc tgctttccat tttgcatttc aaatgtgatt	8220
ttaacttaca ttatacaaagc ctgaaagggtc agtgcaggat ctgctttccat tttgcattt	8280
caaatagtgtat tttaacttacattatacaaa gctgaaaggat cagtcgttgc tctgtgttgc	8340
ttcctataact ttcttagagaa taggaacttc ggaataggaa cttcgttgc tgcaggccc	8400
gggctcgacc agctgtgcgc atagtggttgc gaaatcgataa aataaaagat tttatgttgc	8460
ctccagaaaa aaaaaaaaaat gaaagacccc acctgttaggtt ttggcaagct acgcacaccc	8520
ctcaactcgcc ggcgcaggatcc tccgatttgc tgagtcgttgc gggtaaccgt gtatccaaata	8580
aaccctcttgc cagttgcatttgc cgtactgttgc tctcgcttgc ctttggggagg gtctcttgc	8640
agtgttgcatttgc taccgttgc cgggggttgc tcacatgttgc catgttatcaa aatataatttgc	8700
gttttttttc ttaagtatttt acattaaatgc gccatagtttgc cattatgaa tcggccaaacg	8760
cgccggggaga ggcgggttgc gtattggcg ctcttcgttgc tctcgcttgc ctgactcgct	8820
ggcgctcgatc gttcggttgc ggcggcggttgc atcagcttgc tcaaaaggcgtaatacggtt	8880
atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc	8940
cagggaaacgtt aaaaaggccg cgttgcgttgc gttttccat aggctccccc cccctgtacgt	9000
geatcacaaaa aatcgacgtt caagtcgttgc gtggcgaaac cggacaggac tataaaagata	9060
ccagggcgatcc cccctggaa gtccttcgttgc gtcgttgc tgcgttgc	9120
cgatgttgcatttgc tccgttgc ctttgcgttgc aagcggttgc gtttgcgttgc gtcgttgc	9180
taggttatctc agttcggttgc aggtcggttgc ctccaaatgttgc ggctgtgttgc acgaaccccc	9240
cggttgcgttgc cccctggaa gtccttcgttgc gtcgttgc tgcgttgc	9300
acacgacttca tcggccacttgc cagcggccatc tggtaacagg attagcaggatc cgaggatgt	9360
aggcggttgc acagagtttgc tgaagttgttgc gcctaaacttgc ggctacacta gaaggacgt	9420
atttggtatac tgcgttgc tgaagttgttgc tacatttcgttgc ttgtgtgttgc acgaaccccc	9480
atccggcaaa caaaccaccc ctggatgttgc tgggtttttt gtttgcgttgc acgcgttac	9540
ggcgcaaaaa aaggatctc aagaagatcc tttgtatctt tctacgggttgc acgcgttac	9600
gtggaaacgaa aactcacgtt aaggatctt ggtcatgaga ttatcaaaaa ggatcttac	9660

-continued

```

ctagatcctt ttgcggccgc aaatcaatct aaagtatata ttagtaact tggctgaca 9720
gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtctattt cggtcatcca 9780
tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 9840
ccagtgctgc aatgataaccg cgagacccac gtcacccgc tccagattt tcagcaataa 9900
accagccagc cggaaggggcc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc 9960
agtctattaa ttgttgcgg gaagcttagag taagtagttc gccagttaat agtttgcga 10020
acgttgttgc cattgtgtca ggcacatgtgg tgtaacgttc gtcgttttgtt atgggttcat 10080
tcagctccgg ttcccaacga tcaaggcagag ttacatgtatcc ccccatgttg tgcaaaaaag 10140
cggttagctc ctccggctt ccgatcggtt tcagaagtaa gttggccgca gtgttatcac 10200
tcatgggttat ggcacgactg cataatttctc ttactgtcat gccatccgta agatgtttt 10260
ctgtgactgg ttagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 10320
gtcttgcggc ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc 10380
tcatcatgttgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgtttagat 10440
ccagttcgat gtaaccact cgtgcaccca actgatcttc agcatcttt actttcacca 10500
gggtttctgg gtgagcaaaa acaggaaggc aaaaatgcgc aaaaaaggga ataagggcga 10560
cacggaaatg ttgaataactc atactcttcc tttttcaata ttattgttgcg atttatcagg 10620
gttattgtct catgagcgga tacatatttgc aatgtatttgc gaaaataaaa caaatagggg 10680
ttccgegcac atttccccga aaagtgcac 10710

```

<210> SEQ ID NO 9  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I1

&lt;400&gt; SEQUENCE: 9

cgccctctct tcctccatcc

20

<210> SEQ ID NO 10  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I3

&lt;400&gt; SEQUENCE: 10

actcttccgc ttccctcgctc caccgcggct tcgagaccgt

40

<210> SEQ ID NO 11  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I2

&lt;400&gt; SEQUENCE: 11

gggcctcttc gctattacgc

20

<210> SEQ ID NO 12  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:

-continued

<223> OTHER INFORMATION: primer I4

<400> SEQUENCE: 12

acggtctcga agccgcggtg gagcgaggaa gcggaagagt 40

<210> SEQ ID NO 13  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I6

<400> SEQUENCE: 13

gctcctcgcc cttgctcacc 20

<210> SEQ ID NO 14  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I8

<400> SEQUENCE: 14

tagaagcggt tttcgggaga atacgactca ctataggcg 40

<210> SEQ ID NO 15  
<211> LENGTH: 20  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I5

<400> SEQUENCE: 15

tgcgtggcctt ttgctcacat 20

<210> SEQ ID NO 16  
<211> LENGTH: 40  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer I7

<400> SEQUENCE: 16

cgcctatacg tgagtcgtat tctcccgaaa accgcttcta 40

<210> SEQ ID NO 17  
<211> LENGTH: 71  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide P5

<400> SEQUENCE: 17

gatcctgcac tgaccttca gctttgtata atgtaagtta aaatcacatt tgaaaatgcaa 60

atggaaaagc a 71

<210> SEQ ID NO 18  
<211> LENGTH: 71  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide P6

<400> SEQUENCE: 18

gatctgcttt tccatttgca tttcaaatgt gatttaact tacattatac aaagctgaaa 60

-continued

ggtcagtgca g 71

<210> SEQ ID NO 19  
<211> LENGTH: 45  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer P7

<400> SEQUENCE: 19

gggggctgca gacttacagc ggatccctc aggcacccggg cttgc 45

<210> SEQ ID NO 20  
<211> LENGTH: 36  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer P8

<400> SEQUENCE: 20

gggggctgca gccaatatga ccgagttacaa gccccac 36

<210> SEQ ID NO 21  
<211> LENGTH: 48  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer  
<220> FEATURE:  
<221> NAME/KEY: misc\_feature  
<222> LOCATION: 36..46  
<223> OTHER INFORMATION: n represents inosine at positions 36, 37, 40,  
41, 45 and 46

<400> SEQUENCE: 21

ctactactac taggccacgc gtcgactagt acgggnngn ngggnngg 48

<210> SEQ ID NO 22  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 22

gccagggttt tcccagtac ga 22

<210> SEQ ID NO 23  
<211> LENGTH: 32  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 23

ctactactac taggccacgc gtcgactagt ac 32

<210> SEQ ID NO 24  
<211> LENGTH: 38  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: primer

<400> SEQUENCE: 24

tgtaaaacga cggccagtgt gaaggctgtg cgaggccg 38

-continued

<210> SEQ ID NO 25  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide I16

<400> SEQUENCE: 25

cgagccccag ctggttttt c

21

<210> SEQ ID NO 26  
<211> LENGTH: 25  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide SR1

&lt;400&gt; SEQUENCE: 26

gctagcttgc caaacctaca ggtgg

25

<210> SEQ ID NO 27  
<211> LENGTH: 21  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide I15

&lt;400&gt; SEQUENCE: 27

gtctcagaag ccatagagcc c

21

<210> SEQ ID NO 28  
<211> LENGTH: 24  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide ISR2

&lt;400&gt; SEQUENCE: 28

gccaaaccta caggtgggtt cttt

24

<210> SEQ ID NO 29  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide I14

&lt;400&gt; SEQUENCE: 29

actatcccga ccgccttact gc

22

<210> SEQ ID NO 30  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:  
<223> OTHER INFORMATION: oligonucleotide iPCRu3

&lt;400&gt; SEQUENCE: 30

cctccgattt actgagtcgc cc

22

<210> SEQ ID NO 31  
<211> LENGTH: 22  
<212> TYPE: DNA  
<213> ORGANISM: Artificial  
<220> FEATURE:

-continued

&lt;223&gt; OTHER INFORMATION: oligonucleotide I13

&lt;400&gt; SEQUENCE: 31

tgtttgacc gctgggatct gc

22

&lt;210&gt; SEQ ID NO 32

&lt;211&gt; LENGTH: 22

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: oligonucleotide iPCRu4

&lt;400&gt; SEQUENCE: 32

tacccgtgtta tccaaataaac cc

22

&lt;210&gt; SEQ ID NO 33

&lt;211&gt; LENGTH: 64

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: mus musculus

&lt;400&gt; SEQUENCE: 33

tctttgttgc ttccagcttt gtataagtta gttaaaatca catttgaat gcaaatggaa

60

aagc

64

The invention claimed is:

1. A gene trap vector comprising a gene disruption and selection cassette (GDSC) and one or more non-viral, cell-type-specific enhancer elements that are active in mammalian cells, wherein said GDSC comprises a promoterless reporter gene and/or a promoterless selectable marker gene flanked upstream by a 3' splice acceptor (SA) site and downstream by a transcriptional termination polyA sequence.

2. The gene trap vector of claim 1, wherein the vector is a plasmid.

3. The gene trap vector of claim 1, wherein said one or more non-viral, cell-type-specific enhancer elements (i) are located upstream of the GDSC, and/or (ii) are located downstream of the GDSC.

4. The gene trap vector of claim 1, wherein the one or more non-viral, cell-type-specific enhancer elements contain at least one binding site for a transcription activating factor.

5. The gene trap vector of claim 1, wherein the one or more non-viral, cell-type-specific enhancer elements contain binding sites that bind transcription activation factors in a sequence-specific manner.

6. The gene trap vector of claim 5, wherein the binding sites are arranged as tandem repeats.

7. The gene trap vector of claim 1, wherein the one or more non-viral, cell-type-specific enhancer elements are selected from the group consisting of hormone responsive elements and transcription factor binding elements.

8. The gene trap vector of claim 1, wherein the one or more non-viral, cell-type-specific enhancer elements are tandem repeats that comprise a transcription factor binding site selected from the group consisting of NF-kB, Oct2, Oct3, and Oct4 transcription factor binding sites.

9. The gene trap vector of claim 8, wherein the transcription factor binding sites are the Oct-4 transcription factor binding sites.

10. The gene trap vector of claim 9, wherein the Oct-4 transcription factor binding site are inserted between two homotypic or heterotypic site-specific recombination targets.

30 11. The gene trap vector of claim 1, wherein the reporter gene is a R-galactosidase neomycinphosphotransferase fusion gene.

12. The gene trap vector of claim 1, further comprising a puromycin resistance gene inserted upstream or downstream of the GDSC.

35 13. The gene trap vector of claim 12, wherein the puromycin resistance gene is in direct or inverse transcriptional orientation relative to the GDSC.

40 14. The gene trap vector of claim 1, further comprising site-specific recombinase recognition elements that are oriented in opposing directions and flank the GDSC and the one or more enhancer elements such that inversion of the GDSC and the one or more enhancer element is mediated in the presence of the site-specific recombinase.

45 15. The gene trap vector of claim 14, wherein the recombinase is a Cre or a Flp recombinase.

50 16. The gene trap vector of claim 1, wherein the vector comprises a 5' and a 3' nucleotide sequence homologous to the 5' and the 3' sequence of an intron of a target gene, wherein the nucleotide sequences flank the GDSC and the one or more non-viral, cell-type-specific enhancer elements and further mediate homologous recombination at the intron.

55 17. A mammalian cell comprising the gene trap vector of claim 1.

18. A retroviral gene trap vector comprising a GDSC and one or more cell-type-specific enhancer elements that are active in mammalian cells, wherein said GDSC comprises a promoterless reporter gene and/or a promoterless selectable marker gene flanked upstream by a 3' splice acceptor (SA) site and downstream by a transcriptional termination polyA sequence.

60 19. The retroviral gene trap vector of claim 18, which is eFlip3ROSAgeo or eFlip6ROSAgeo having the nucleotide sequence of SEQ ID Nos:4 and 5, respectively.

\* \* \* \* \*